THIS WEEK

EDITORIALS

SLOPPY SCIENCE The perils and pitfalls of work with big data sets **p.406**

WORLD VIEW Jargon helps language to make sense of the world **p.407**

PESTS Cane-toad toxin could help Australia to battle back **p.409**

Secret disservice

Staff-surveillance efforts by government agencies must not contravene the rights of whistle-blowers, as the US Food and Drug Administration is accused of doing.

With supreme irony, a clandestine effort by the US Food and Drug Administration (FDA) to spy on its own staff was exposed when some 80,000 pages of documents gathered during the operation were accidentally published on the Internet.

Using monitoring software, the agency collected the communications of five staff scientists whom it suspected of going public with concerns over the safety of imaging devices designed to identify breast or colon cancers that had been approved or were in line for approval by the agency (see page 418). Four of the scientists lost their jobs.

Beginning in the spring of 2010, the spying operation tracked all communications to and from the scientists' government-issued computers, capturing communications with their lawyers, each other, at least one member of Congress, congressional staff members and the media. The exercise allowed the agency to compile a list of 21 'actors' — including four further agency scientists — who an unidentified FDA official wrote were engaged in a "collaborative plan" to "defame" the agency.

According to a lawsuit against the government filed by the scientists, the operation captured private e-mails sent on personal time, on private networks and on private equipment, and private e-mails sent to or received from other private e-mail accounts. According to US Senator Charles Grassley (Republican, Iowa), who is investigating the affair, the entire operation was authorized by the head lawyer at the FDA.

The FDA notes, rightly, that the federal Food, Drug, and Cosmetic Act makes it illegal for the agency's employees to release confidential commercial information without legal authorization. It adds that it "had evidence suggesting" that "a small number of FDA employees … might be responsible for the unauthorized disclosure of proprietary information". The operation, it says, "was only intended to identify the source of the unauthorized disclosures, if possible and to identify any further unauthorized disclosures".

Only in court will the FDA get to tell its side of the story fully. But in a world in which surveillance software can capture communications by keyword, the extent of the FDA's operation is breathtakingly broad and intrusive. According to the lawsuit, it captured privileged communications between the scientists and their lawyers, an official complaint the scientists sent to the government's Office of Special Counsel — which investigates complaints made by whistle-blowers — and their correspondence with the Department of Justice and the inspector-general for the Department of Health and Human Services (DHHS), the FDA's parent agency. The inspector-general is tasked with maintaining the integrity of DHHS programmes by rooting out waste, fraud and abuse.

Disturbingly, the FDA ignored consistent findings from the inspector-general. Asked by the FDA to investigate the scientists it suspected in mid-May 2010, the inspector-general quickly determined that the FDA lacked evidence of any criminal conduct and declined to take action.

The inspector-general noted, according to the lawsuit, that the disclosures the agency alleged against the scientists were a protected whistle-blower activity. At that point, the FDA was monitoring only one of the scientists' computers. Instead of ceasing the surveillance, within weeks it installed spyware on the other four scientists' computers, and asked the inspector-general to re-investigate. The inspector-general again demurred, noting, according to the lawsuit, that the Department of Justice had also declined to prosecute. The Office of Special Counsel has been so alarmed by the FDA's behaviour that it issued a government-wide warning last month, reminding agencies that any surveillance they conduct must not trample on the rights of whistle-blowers.

"The FDA's operation is breathtakingly broad and intrusive."

The whole sorry saga is made worse by the fact that it centres on the public's health, and the efforts of civil servants to protect it. In the most charitable construction, the agency's leadership made a glaring but legal error in judgement. But the available evidence seems rather to point to egregious, chilling and very possibly illegal conduct by the FDA.

There is one thing that can be done, and quickly, to mitigate the unquestionable discomfort this case will cause any whistle-blower who wants to report bad behaviour by government colleagues. In May, the US Senate unanimously passed a bill to significantly enhance legal protections for federal workers, including scientists, who blow the whistle — protections that have been seriously eroded by years of bad court decisions. The House, which must now pass the bill for it to become law, should move quickly to do so, especially because this is one of the few issues on which law-makers on both sides have been able to agree. ■

Protect and serve

A 'health check' of protected ecological areas reveals an alarming decline in biodiversity.

Protecting designated areas of ecological value is one of the most popular conservation tools for safeguarding biodiversity. As deforestation advances, the theory goes, these protected areas offer sanctuary for threatened species and natural ecosystem processes. By 2020, the 193 nations that have signed up to the Convention on Biological Diversity — the United States is not among them — hope to protect at least 17% of the planet's biodiversity-rich areas in this way. But is drawing artificial lines around ecologically valuable land an effective method for protecting biodiversity?

As a conservation mechanism, protected areas have a chequered history, and have faced particular criticism because some exclude poor local populations who want to gather food, wood and other resources from the forests on which they depend. Management of the thousands

of protected areas varies around the world, and biodiversity scientists have long suspected that areas with better safeguards to stop illegal encroachment for logging, agriculture or other activities are more likely to be a safe haven for biodiversity. However, demonstrating that such exclusion is effective is not straightforward.

Assessments of ecosystems and biodiversity are hampered by piecemeal data collection that uses incomparable methodologies, and the data have large gaps. That makes it difficult to draw broad conclusions about the global health of biodiversity, which many scientists say is a key reason why biodiversity has failed to climb political agendas, even as extinctions of animals and plants continue to rise. Around one-fifth of all 5,490 described mammal species are at risk of extinction, according to data from the International Union for Conservation of Nature.

These difficulties are driven by funding constraints and the intricacies of monitoring and assessing complex ecosystems. Nevertheless, some informative evaluations do exist, including the biennial Living Planet Index drawn up by the conservation group WWF, which follows trends in populations of species around the world to give indications of the state of global biological diversity.

More of these data gaps have now been filled, thanks to an assessment of protected tropical forest areas led by William Laurance, a conservation biologist at James Cook University in Cairns, Australia, which offers a snapshot of global biodiversity and reviews the success of this tool.

In a Herculean effort, Laurance and his team, along with more than 100 co-authors, systematically collected standardized data on environmental changes over the past 20–30 years in 60 protected areas across the world's major tropical regions of Africa, America and Asia. The data include changes in the abundance of 31 groups of species, including primates, freshwater fish and exotic plants, and 21 potential drivers of environmental change, such as road building and hunting.

To gather the information, the team conducted 262 interviews with field biologists and environmental scientists, and asked them each to complete a detailed 10-page questionnaire. It took the team around four years to gather the data, but Laurance reckons that to do so from scratch would cost billions of dollars and take 20–30 years. The effort paid off, and the results appear online on *Nature*'s website this week (W. Laurance *et al. Nature* http://dx.doi.org/10.1038/nature11318; 2012).

"Around half of the reserves are experiencing a severe loss of biodiversity."

The team found that around half of the reserves are experiencing a severe loss of biodiversity. Crucially, the researchers also demonstrated the negative impact that environmental changes immediately outside the areas have on the health of the habitat inside the reserves. And they found that destructive activities such as forest clearance, fires and logging increasingly reach up to the edge of the protected areas. The results show that 85% of the reserves experienced declines in surrounding forest cover over the study period, whereas only 2% gained forest. These findings will not surprise biodiversity scientists, who have long been aware of such trends, but now have the data to show it.

The work highlights, yet again, the growing challenges that threaten the success of the protected-reserves model. Many reserves simply do not function as intended, but there are few alternative approaches. The results underscore the importance of better management of areas around the reserves. Rather than treating these ecological havens as islands, Laurance and his team recommend implementing buffer zones around protected areas to cushion the blow. The reserves can protect wildlife, but we must first protect the reserves. ■

Error prone

Biologists must realize the pitfalls of work on massive amounts of data.

Genomics has the potential to revolutionize medical care, but it is becoming increasingly clear that the field is having to deal with growing pains.

In a Comment piece this week, Daniel MacArthur, a researcher at Massachusetts General Hospital in Boston, argues that the massive pools of data generated in even routine genome studies make it easy to misinterpret artefacts as biologically important results (see page 427). Such false positives, he says, can lead to embarrassing retractions, futile projects and stalled careers. More careful attention to methods and greater awareness of the potential pitfalls will help to cut down on the needless mistakes.

In a field as competitive as genomics, scientists will inevitably seek faster, more efficient ways to generate and analyse data. Just this week, the firm Ion Torrent in Guilford, Connecticut — part of Life Technologies in Carlsbad, California — announced that it will tackle a competition to accurately sequence 100 genomes in 30 days for less than US$1,000 per genome — and to win the US$10-million prize offered by the X Prize Foundation in Playa Vista, California (see page 417).

Genomics is not the only field of science to battle with quality-control issues. In March, *Nature* lamented the high number of corrections to research papers in the life sciences that arise from avoidable errors (see *Nature* **483,** 509; 2012). Scientists are making too many careless mistakes, and those mistakes are getting published.

Much of this sloppy science comes from the pressure to generate 'surprising' results and to publish them quickly, even though they are more likely to be driven by errors than are findings that more or less follow from previous work. A researcher who reveals something exciting is more likely to get a high-profile paper (and a permanent position) than is someone who spends years providing solid evidence for something that everyone in the field expected to be true.

This pressure extends throughout the careers of scientists, and is compounded by the preference of journals (including *Nature*) to publish significant findings — and of the media to report them. MacArthur asks scientists to weigh up the importance of avoiding being scooped against the embarrassment of a mistake, but to an ambitious scientist in a competitive field such as genomics, the risk of being out-published will often outweigh the potential damage of retraction.

Many areas of the life sciences now work with massive amounts of data, so technology-based artefacts are unlikely to be restricted to genomics. Any life scientist who works at a university or is affiliated with a hospital can now collect human samples and sequence them to create huge amounts of genomic data, with which they are perhaps not used to working. The problem goes beyond analysis — time and time again, biologists fail to design experiments properly, and so submit underpowered studies that have an insufficient sample size and trumpet chance observations as biological effects.

The problems are not hard to solve. Biologists must seek relevant training in experimental methods and collaborate with good statisticians. Principal investigators have a responsibility to their labs and to colleagues to ensure that any data they publish are robust. And the efforts of peer reviewers who thoroughly reanalyse data to double-check that submissions are solid deserve more formal acknowledgement, albeit in private.

Meanwhile, researchers who deal with large amounts of data must agree on standards that will protect against avoidable errors. Fields such as RNA sequencing have been slow to establish such guidelines (see *Nature* **484,** 428; 2012), but others have shown that it can be done. The human-genetics community, for instance, has established criteria for genome-wide association studies to ensure that findings are rigorous and comparable. Less-proactive genomics fields, and the rest of biology, should follow that lead. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

WORLD VIEW *A personal take on events*

Writers should not fear jargon

Researchers use complex language for a specific purpose, and science writers should be clear about what those reasons are, says **Trevor Quirk**.

Who needs jargon? Last month a physics PhD student at the University of Innsbruck, Austria, won a competition to explain the concept of a flame in words that an 11-year-old could understand. Ben Ames, the winner, made a 7.5-minute video, which introduced words such as 'oxidation' and 'pyrolysis', only to parody them.

The very premise of this contest speaks to the aversion we science writers have for jargon. Many seem to assume that the pompous, sterile language of scientific literature has been designed to prevent our understanding it. Reading the stuff seems a kind of sadistic chore. Translating it? Unspeakable. So you can imagine the unpopularity of my belief that jargon is not only integral to scientific discourse but also has a place in public discussion.

Certainly, there is a lot in academic writing that I really can't defend — needless passive phrasing, for instance — but I also think a flip rejection of jargon reflects a greater hostility towards difficult language that pervades modern culture.

When faced with any jargon — scientific, business-speak, legalese — people tend to presume that every term could be substituted with something more colloquial. At first, it might seem unnecessary for economists to use the French word 'tranche' instead of 'layer', 'slice' or 'cut'. But common synonyms are problematic because they can be swapped and easily confused for each other.

Specialized terms capture the complexity and specificity of scientific concepts. Consider astronomy, in which both 'photometry' and 'spectroscopy' denote techniques that could be described in a jargon-free way as 'methods of studying light'. Yet photometry is the measurement of light's intensity and spectroscopy is the study of its relationship to its source. Both are complex, important and highly specific techniques. No other words in the English language encapsulate their meaning quite as well, and if they are dismissed as jargon, then that meaning is lost.

Scientific literature abounds with distinctions that can seem pedantic. Consider the 'intrinsically photosensitive retinal ganglion cell' — or 'ipRGC'. The term refers to a specific type of neuron located in the eye, and although the phrase is no fun to parse, every word in it is important. A 'ganglion', loosely defined, is a mass of tissue, often found in the eye, so 'cell' refers to a specific part of that tissue. Not all ganglia are found in the retina, thus 'retinal' is justified. And not all retinal ganglia are 'intrinsically photosensitive', so that stays, too. This is perhaps the hardest truth for the more idealistic science writers to swallow. It would take paragraphs of explanation to make all of the other scientific distinctions contained in the term 'ipRGC'. Many science writers would hack away at the term (they call this process 'distilling'), finally calling it, perhaps, a 'special kind of ganglion' or a 'neuron located in the eye'. Such wording is easier to understand but it does not present the whole truth. I am not arguing that science writers should always use jargon, but I do want to point out what can be lost when they do not.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/bmgfqy

The truth tends to be complicated, and here jargon offers its most obvious perk: compression. There is emotional compression in much writing, perhaps best seen in this (perhaps apocryphal) work by Ernest Hemingway: "For sale: baby shoes, never worn." Technical writers use jargon to compress information. A reluctance to use and engage with it can have serious consequences. Consider terms such as 'credit default swap' — there is a whole backwards school of thought that suggests that these terms were designed simply to confuse and bore people into apathy and inaction. To me, this seems like an oblique justification for not caring enough, and highlights a general reluctance to labour for meaning.

A FLIP REJECTION OF JARGON REFLECTS A GREATER HOSTILITY TO DIFFICULT LANGUAGE.

Jargon requires work from a general readership. But it also requires work from those who use it. Organic and physical chemists speak entirely different languages, as do extragalactic and stellar astronomers, and glaciologists and hydrologists. These linguistic divisions are not created out of the desire to alienate with lofty and overcomplicated language, they are a natural consequence of getting at the unthinkable complexity of the natural universe. To this purpose, jargon is a necessity, as is the labour required to understand it.

Other words are just as labour intensive as jargon. It takes real work to understand the meanings of words such as 'portentous' and 'pretentious' or 'voracious' and 'veracious'; or to make the small but meaningful distinction between 'impel' and 'compel'.

I find it troubling that the same antipathy that some writers express towards jargon has taken root in the public's general attitude towards erudite language. I submit that this is no coincidence. People seem to resent not just specialized language, but any language that requires a large degree of labour to understand, appreciate and use. When hearing someone complaining of having to consult a dictionary — especially when that consultation does not even involve moving from the computer in front of them — I am overcome with the desire to grab that person's lapels and shake them until their teeth rattle. Why are people so unwilling to work for the pleasures and insights that language harbours? When writers avoid jargon unquestioningly, readers start to think that it serves no purpose. The world increases in complexity every day, and we should not let shrink our capacity to describe it. ■

Trevor Quirk *is a science writer who is interning at* Harper's Magazine.
e-mail: tsquirk@bu.edu

RESEARCH HIGHLIGHTS

Selections from the scientific literature

MATERIALS

Rolling out data storage

The data-storage capacity of CDs and DVDs is constrained by their size, so a group of researchers has devised a continuous process that makes long sheets of a many-layered optical storage medium.

Kenneth Singer of Case Western Reserve University in Cleveland, Ohio, and his colleagues began by melting and stacking two polymer layers and then pushing them through a series of 'multipliers', each of which doubles the number of layers. After the multi-layer melt was spread into a thin but dense optical-storage film, it could be cut and formed into various shapes and sizes. The team then used a laser to write data on each of 23 layers.

The group hopes that the method will eventually produce low-cost optical data film hundreds of metres long that can store terabytes or even petabytes of data — roughly equivalent to one million DVDs.

Adv. Mater. **http://dx.doi.org/10.1002/adma.201200669 (2012)**

CANCER IMMUNOLOGY

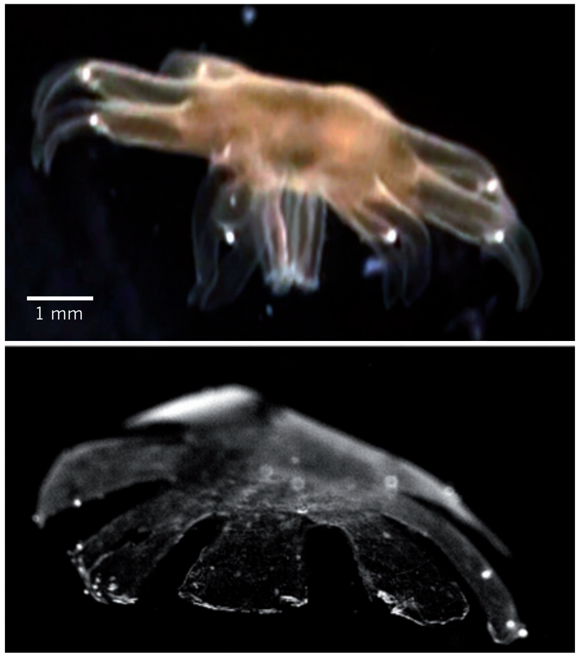
Pathway from breast to bone

The likelihood that breast tumours will spread to the bone — a potentially deadly process — is enhanced when an immune-signalling pathway in the cancer cells is blocked.

Belinda Parker at the Peter MacCallum Cancer Centre in Melbourne, Paul Hertzog at Monash University in Clayton, both in Australia, and their team compared gene expression in cells from primary mouse breast tumours with those from bone metastases. They found that in the metastases, many genes sharing the same pathways as a protein called IRF7 were suppressed. This protein regulates the body's responses to an immune-signalling molecule called interferon, a version of which is used in the treatment of certain cancers. Restoring IRF7 signalling in tumour cells reduced bone metastases and boosted immune activity and metastasis-free survival time.

Patients with breast cancer and low expression levels of IRF7-associated genes in their primary tumours could be at higher risk of metastasis, the authors say.

Nature Med. **http://dx.doi.org/10.1038/nm.2830 (2012)**



J. NAWROTH

BIOENGINEERING

Engineered 'jellyfish'

By combining rat tissue and silicone, researchers have engineered a device that mimics the movements of jellyfish.

Kevin Kit Parker at Harvard University in Cambridge, Massachusetts, John Dabiri at the California Institute of Technology in Pasadena and their colleagues used computer imaging and tissue experiments to characterize the muscular structure of jellyfish (pictured top) and the motion — a quick contraction and slow recoil — that the animals use to feed and to propel themselves through water. By growing rat heart-muscle cells on a flexible silicone frame, the team then built a pump (bottom) capable of mimicking jellyfish shape, movement and fluid dynamics. Applying an electrical field to a water bath containing the 'jellyfish' caused it to contract with a power and at a rate similar to those of natural jellyfish.

The researchers suggest that the same design strategy could be used to produce other synthetic muscular pumps or models of simple organisms.

Nature Biotechnol. **http://dx.doi.org/10.1038/nbt.2269 (2012)**

For a longer story on this research, see go.nature.com/xfr6jc

MICROBIOLOGY

Recycling at root of arsenic 'life'

The reported discovery of a bacterial strain able to use arsenic in place of phosphorus to make essential molecules such as DNA generated much controversy and has recently been refuted. Murray Deutscher and his colleagues at the University of Miami in Florida provide an alternative explanation for the puzzling observation that bacteria of the strain *Halomonadaceae* GFAJ-1 grow when supplied with arsenate instead of phosphate.

The researchers used a radioactive isotope to label ribosomes — the cell's protein-making machinery — in the bacterium *Escherichia coli*. They found that arsenate causes ribosomes to degrade, a process that releases phosphate. After being cultured for about 80 hours in media that contained arsenate but not phosphate, a few arsenate-tolerant *E. coli* cells began to grow, slowly. Rather than growing by replacing phosphorus with arsenic, these cells recycled phosphorus released by deteriorating ribosomes, the authors suggest.

J. Biol. Chem. **http://dx.doi.org/10.1074/jbc.C112.394403 (2012)**

HIV

Human response in model mice

Mice that carry several key parts of the human immune system replicate the human immune response to HIV. The model could be used to test HIV vaccine candidates — an effort that has been hindered by the lack of a suitable small-animal model.

'BLT' mice are engrafted with human bone-marrow cells, liver and thymus tissue, and produce a functional human immune system, including a robust population of human $CD4^+$ T cells, the target of HIV. Todd Allen of the Ragon Institute in Charlestown, Massachusetts, and his colleagues show that, during the acute phase of infection, these mice generate HIV-specific killer T cell responses that closely resemble those in humans. Importantly, the virus rapidly evolves to escape these responses, just as it does in humans. What's more, BLT mice that express HLA-B57, a human gene variant that protects against HIV, show better control of HIV replication — just as humans with this variant do.

Sci. Transl. Med. **4, 143ra98 (2012)**

ANIMAL BEHAVIOUR

Squid 'fly' faster than they swim

Squid can exhibit brief periods of 'flight' above water, perhaps because this could be a more efficient form of movement than swimming.

The animals propel themselves through the air (**pictured**) by forcing water out of their mantles. Ron O'Dor at Dalhousie University in Halifax, Canada, and his colleagues examined laboratory and field data on four species of squid in air and water. They found that velocities in the air could reach 37 body lengths per second, whereas the maximum seen in water was a mere 11 body lengths. Acceleration was also significantly higher in air than in water.

The team suggests that the advantages of squid 'flight' might mean that it is more common than currently thought, and that it may even reduce the cost of long-distance migrations.

Deep-Sea Res. Pt II **http://dx.doi.org/10.1016/j.dsr2.2012.07.002 (2012)**

A. PIERCE/SPLASHDOWNDIRECT/REX FEATURES

GEOLOGY

Greenland's ancient impact

A circular region more than 100 kilometres in diameter on the southwest coast of Greenland bears the footprint of a massive impact that occurred about 3 billion years ago.

Adam Garde of the Geological Survey of Denmark and Greenland in Copenhagen and his colleagues analysed evidence of intense crushing and heating in a formation whose features would have been buried about 25 kilometres below Earth's surface at the time of impact. In a central area 35–50 kilometres in diameter, all pre-existing rock structures were destroyed and partly melted. The authors also found evidence of fracturing, deformation and some melting within a diameter of 100–140 kilometres.

Large impacts were common early in Earth's history, but much of the evidence has been eroded by climate and tectonics. The authors suggest that the Maniitsoq structure, which they named after the nearest town, is the oldest and largest impact structure on record.

Earth Planet. Sci. Lett. **http://dx.doi.org/10.1016/j.epsl.2012.04.026 (2012)**

COMMUNITY CHOICE

The most viewed papers in science

ECOLOGY

Toxins for cane-toad control

HIGHLY READ on rspb.royalsocietypublishing.org in June.

The invasive cane toad is wreaking havoc in Australian tropical ecosystems. But the toxin that this animal produces could be used to control its populations.

Cane toad (*Rhinella marina*; pictured) tadpoles seek out and consume newly laid eggs of their own species, to reduce competition for resources. To identify the chemical cues that drive this behaviour, Richard Shine of the University of Sydney in Australia and his colleagues analysed secretions from cane toad eggs and adults, and found that bufadienolide toxins dominate both. Using toxin samples as bait, the researchers set funnel traps in two floodplain ponds. Cane tadpoles made up more than 98% of the trapped animals. Laboratory trials with native species and cane tadpoles showed that the toxin repels native tadpoles.

Proc. R. Soc. B **http://dx.doi.org/10.1098/rspb.2012.0821 (2012)**

P. NASKRECKI/MINDEN PICTURES/FLPA

CANCER

New target for melanoma fight

Melanoma, the most lethal form of skin cancer, is difficult to treat, but researchers have identified a potential drug target: a protein that blocks the effect of the tumour suppressor p53.

Jean-Christophe Marine at the Dutch-speaking Catholic University of Leuven in Belgium and his colleagues found that levels of the protein, MDM4, are elevated in about 65% of human melanomas. When they overexpressed MDM4 in melanoma-prone mice, the researchers found that nearly all the animals developed melanoma, compared with only about half of the control mice, and the animals did so much more quickly. Treating human melanoma cells with a molecule that blocks the interaction between MDM4 and p53 restored p53 activity, with those cells undergoing apoptosis, or programmed cell death. This treatment also rendered melanoma cells more sensitive to two common chemotherapy drugs.

Inhibiting MDM4 could boost the effectiveness of other drugs, the authors suggest.

Nature Med. **http://dx.doi.org/10.1038/nm.2863 (2012)**

CORRECTION

In the Research Highlight 'Hormone linked to depression' (*Nature* **487,** 274; 2012), Xin-Yun Lu's affiliation should have been given as the University of Texas Health Science Center at San Antonio, not the University of Texas at San Antonio.

NATURE.COM

For the latest research published by *Nature* visit:

www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Climategate closed

Regional police have closed their investigation into the November 2009 release of e-mails from scientists at the Climatic Research Unit (CRU) at the University of East Anglia (UEA) in Norwich, UK. The Norfolk constabulary said on 18 July that they had no realistic prospect of identifying the offenders within legal time constraints, but that the data breach “was the result of a sophisticated and carefully orchestrated attack on the CRU’s data files, carried out remotely via the Internet”. The investigation cleared anyone associated with UEA from involvement with the crime. See go.nature.com/rzkbwg for more.

S. MCCALL/GETTY IMAGES

US crops wilt in drought

The most extensive US drought in more than half a century has seen prices for maize (corn) and soya beans leap to record highs. The National Climatic Data Center in Asheville, North Carolina, said on 16 July that moderate to extreme drought was affecting 55% of the continental United States — the highest proportion since December 1956. On 23 July, the US Department of Agriculture said that it would help farmers and ranchers by allowing haying and grazing on some land usually protected in conservation programmes.

Romania plagiarism

Two investigations into the case of alleged plagiarism by Romania’s prime minister, Victor Ponta, have reached opposite conclusions, increasing the tension in a fierce struggle over political power in Bucharest. Ponta is accused of copying large sections of his 2003 PhD thesis on the International Criminal Court (see *Nature* **486,** 305; 2012). On 19 July, the government-appointed National Ethics Council rejected the plagiarism charges against Ponta. But one day later, an ethics commission set up by the University of Bucharest — which awarded Ponta his PhD — said that he did copy and paste large sections of his thesis. See go.nature.com/mfm8x8 for more.

JOURNALISM AWARD

Freelance journalist Stephen S. Hall last week won the American Geophysical Union’s Walter Sullivan award for Excellence in Science Journalism–Features for his article ‘At fault?’ (*Nature* **477,** 264–269; 2011), which examined why seismologists were placed on trial after an earthquake devastated L’Aquila, Italy, in 2009.

Science education

US President Barack Obama announced a plan on 17 July to create a corps of exceptional, or ‘master’, teachers in science, technology, engineering and mathematics (STEM). Using $1 billion dedicated for the purpose in Obama’s 2013 budget request, the corps would start with 2,500 teachers — rising to 10,000 in the next four years — who would receive an annual bonus of up to $20,000 for their work. The budget request may not pass Congress, but $100 million from an existing teacher-incentive fund is now being used to help schools to identify and develop effective STEM teachers.

Clinical trials

The European Commission has adopted proposals for new rules to replace its directive on clinical trials, which set out how medical research is regulated across Europe. Many scientists criticize the decade-old directive for being excessively bureaucratic, and think that it has driven clinical research away from Europe. The reforms, formally proposed on 17 July, would create a centralized applications procedure for trials that are run in more than one country, and include a plan to inspect the regulation of trials in other countries such as China and India. See go.nature.com/fhslei for more.

EVENTS

Arctic iceberg

A 120-square-kilometre floating tongue of ice broke off the Petermann glacier on the Greenland ice sheet on 16 July. Although almost twice the area of Manhattan, it is less than half the size of the 260-square-kilometre iceberg that broke free of the same glacier (one of Greenland’s largest) in August 2010.

Fukushima work

Workers on 18 July removed two unused fuel rods from the spent fuel pool of Fukushima Daiichi’s unit 4 reactor. The reactor was shut down at the time of the earthquake and tsunami that led to meltdowns in other reactors last March, but the stability of its fuel pool — holding fresh as well as spent fuel rods — has

caused concern. The rods will reportedly be checked for signs of corrosion and moved to another building on site. A more permanent structure will be built over unit 4 before fuel removal can begin in earnest next year.

RESEARCH

Child study revamp

Leaders of the US National Children's Study have released a revised design for the ambitious effort to track influences on the health of 100,000 American children from before birth to age 21. An earlier proposed revamp, designed to cut costs, had been criticized for its selective sampling method; the new proposal, released on 24 July, would enrol women from birthing hospitals across the country, creating a sample representative of the US population. To cut costs, it would track only some 60% of participants while they are in the womb. The US Congress may require a report from the Institute of Medicine before any changes are made to the current design.

PEOPLE

Sally Ride dies

R. SACHS/CNP/CORBIS

Scientist, astronaut and educator Sally Ride (**pictured**), who was the first US woman in space, died on 23 July, aged 61. Ride's first space flight was aboard the shuttle *Challenger* in June 1983; she later served on the commissions investigating the *Challenger* and *Columbia* shuttle disasters. After working at NASA, she headed the California Space Institute in San Diego, and founded a company, Sally Ride Science, to encourage students to study the sciences. See go.nature.com/immsvq for more.

BUSINESS

Gene therapy

Europe's drugs regulator has for the first time recommended a gene therapy for approval. Glybera, a treatment for a rare disease in which patients cannot produce enough lipoprotein lipase (an enzyme crucial for breaking down fat), was backed by the European Medicines Agency's Committee for Medicinal Products for Human Use. The 20 July recommendation must also be endorsed by the European Commission before the therapy becomes available, but a rejection would be unusual. The treatment is owned by uniQure, a company based in Amsterdam; it has been tested on only 27 patients. See go.nature.com/ieekkp for more.

Sequencing prize

A US$10-million contest to sequence the genomes of 100 centenarians in 30 days has its first entrant: Ion Torrent (a subsidiary of Life Technologies). The company, based in Guilford, Connecticut, said on 23 July that it would attempt to win the Archon Genomics X Prize, which is offered by the non-profit X Prize Foundation in Playa Vista, California. The contest requires a cost-per-genome below $1,000. See page 417 for more.

Second obesity pill

Just weeks after it approved its first weight-loss drug in 13 years, the US Food and Drug Administration on 17 July gave the green light to a second obesity treatment: Qsymia (phentermine plus topiramate) from Vivus of Mountain View, California. Just like the recently approved Belviq (lorcaserin), Qsymia was rejected by the agency in 2010 because of safety concerns. New data led to its approval — although it carries a warning that those taking the drug should have their heart rates monitored. See go.nature.com/vtcifa for more.

Gene patents redux

A court ruling that allows companies to patent DNA was challenged again in a Washington DC appeals court hearing on 20 July. A court decided last year that patents on genes are valid, as part of its ruling on patents held by Myriad Genetics of Salt Lake City, Utah. But the case was reopened after two separate patents covering a way to determine drug dosage, from Prometheus Laboratories in San Diego, California, were overturned because they were based on 'laws of nature' (which cannot be patented in the United States). See go.nature.com/8pklwr for more.

COMING UP

29 JULY–3 AUGUST
Europe's largest congress on plant science takes place in Freiburg, Germany; topics include the consequences of climate change, and genetically modifying plants to feed the world's population.
go.nature.com/m128l2

2 AUGUST
The governing board of the Cancer Prevention and Research Institute of Texas meets publicly in Austin. Its agenda includes approval decisions on seven research grants, worth US$39 million, that were sidelined in March amid controversy over the institute's review process (see *Nature* **486,** 169–171; 2012).
go.nature.com/qzpdov

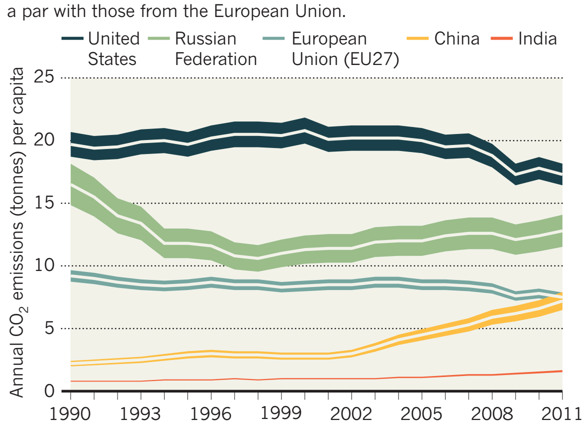
TREND WATCH

SOURCE: JRC/PBL

For years, China has played down concerns about its rising carbon dioxide emissions, saying that on a per-capita basis they are lower than those of their counterparts in the industrialized world. But according to an analysis of 2011 emissions, released on 18 July by the European Commission's Joint Research Centre and the PBL Netherlands Environmental Assessment Agency, this is no longer true (see chart). Total emissions rose 3% in 2011 to around 34 billion tonnes. See go.nature.com/ouxnfg for more.

CHINA'S EMISSIONS RISE

China's per-capita carbon emissions are now on a par with those from the European Union.



NATURE.COM
For daily news updates see:
www.nature.com/news

for TB Drug Development (TB Alliance), a public–private partnership in New York City that sponsored the study on the latest combination therapy.

Spigelman's group wants to speed the development of such treatments. "After a new drug reaches the market, the traditional route is to work from the standard of care and substitute one drug for another. But if you do it in an incremental way, that will give you incrementally better results," Spigelman says. "Our goal is to shorten the path to getting a new, cohesive regimen out there in one fell swoop."

The new combination therapy consists of PA-824, a novel drug in the TB pipeline; moxifloxacin, an antibiotic for treating pneumonia and sinus infections; and pyrazinamide, a TB drug developed in 1952. During a two-week trial, the combination eliminated more than 99% of *M. tuberculosis* in patients' sputum. It promises to be far less taxing for patients than the current regimen for drug-resistant TB, which entails injections and two years of swallowing as many as 20 tablets a day with side effects that can include vomiting, seizures, painful nerve damage and permanent hearing loss. The results, described at the AIDS meeting, were subsequently published on 23 July (A. H. Diacon *et al. Lancet* http://doi.org/h34; 2012).

If full clinical trials confirm its promise, the therapy could prove a boon for HIV-infected people. TB, which is common in its latent form in many parts of the developing world, often becomes active in those who are HIV-positive because of their weakened immune systems.

"When I worked in West Africa," says John Farley, deputy director of the antimicrobial product division at the FDA, "there were times when our HIV patients would die of drug-resistant TB before we could even get them their antiretroviral therapy." Unlike some existing therapies, the new combination does not seem to interact adversely with drugs for HIV.

The therapy was discovered during a broader search for untested permutations of emerging and existing drugs that the TB Alliance began in 2007 with funds from the Bill & Melinda Gates Foundation, the US Agency for International Development, Irish Aid and UK Aid. Its developers hope that the FDA's new guidelines for combination therapies — reserved for emergencies in which patients are dying for lack of treatments — will allow speedy approval. In the past, the FDA requested data to ensure the safety of each independent component of a combination therapy before approving it, says Farley. "But now we think there are other ways to answer that question, through studies in animal models and innovative human trial designs."

The FDA rule change is new enough that most drug developers are still focusing on individual compounds, but that quest has made progress too (see 'Renewed attack on TB'). One drug, delamanid from Otsuka Pharmaceutical based in Tokyo, is under review at the European Medicines Agency. And on 2 July, Janssen Pharmaceuticals, headquartered in Raritan, New Jersey, applied for FDA approval of its TB drug, bedaquiline.

Sequella, a biotechnology firm in Rockville, Maryland, is testing combinations of its drug SQ109 and bedaquiline in mice in anticipation of Janssen's drug getting approval. But Sequella's executive vice-president for corporate development, Alan Klein, says that the company is filing an independent application for its new drug because the FDA's process for combinatorial therapies is simply too new. "It can be risky for both parties because if you don't show good results in combination studies, it might impact your ability to get approval alone," he says.

Like many of the 21,000 attendees at the AIDS meeting, the TB Alliance wants donors to get the message that years of research will soon pay off in lives saved — as long as the money flows. "If we can get the resources to follow this study up," Spigelman vows, "we'll get this combination out." ■

RENEWED ATTACK ON TB

After a long gap, several drugs and a combination therapy for tuberculosis are moving through the pipeline.

Drug	Developer	Clinical phase	Class
Delamanid	Otsuka Pharmaceutical	III (filed for approval for drug-resistant TB)	Nitroimidazole
Bedaquiline	Janssen Pharmaceuticals	II (filed for approval for drug-resistant TB)	Diarylquinoline
SQ109	Sequella	II	Diamine
PA-824	TB Alliance	II	Nitroimidazole
Sutezolid	Pfizer	II	Oxazolidinone
AZD5847	AstraZeneca	II	Oxazolidinone
Combination: PA-824, moxifloxacin, pyrazinamide	TB Alliance	II	Nitroimidazole, fluoroquinolone, nicotinic acid derivative

Source: *2012 Pipeline Report* (iBase/TAG, 2012).

FUNDING

Cuts loom for US science

As budget bills line up, agencies anticipate post-election panic.

BY IVAN SEMENIUK AND HELEN THOMPSON

In an ordinary year, a flat budget for the US National Institutes of Health (NIH) would be considered dire news. This year, it is far from the worst possible outcome. Hanging over the effective decrease in support proposed by the House of Representatives last week is the 'sequester', a pre-programmed budget cut that research advocates say would starve US science-funding agencies.

A sharply divided Congress is showing few signs that it can defuse the situation before the self-imposed fiscal time bomb explodes, in less than six months' time. And even if Congress does manage to introduce last-minute legislation, as many observers expect, the sequester will have cast a shadow over the contentious process of funding science in a time of fiscal constraints — and in an election year.

The sweeping cut, scheduled to take effect on 2 January, is a by-product of last year's Budget Control Act, which requires law-makers to find ways to reduce the federal deficit (see *Nature* **476,** 133-134; 2011). When a congressional committee failed in its remit to do just that last November, the clock began ticking towards an automatic cut that will claw the required amount from across the federal government, including all military and non-military spending that is not required by law. The precise amount to be cut depends on several variables, including tax revenue, but an estimate by the Congressional Budget Office puts it at 7.8% in 2013 for the non-military component.

"Nobody wants to see the sequester, because it's a terrible budgetary tool," says Mike Lubell, director of public affairs for the American Physical Society in Washington DC. "You don't

just take a meat axe and chop off one finger from every pair of hands."

Although no one doubts that the sequester would have a major impact on US science funding (see 'Winners to losers'), no agencies have announced what they will do if it happens. This is mainly because of ambiguity in how the sequester is supposed to be applied: the cuts are described as across-the-board, but how much latitude agencies will have to manage reductions at the programme level remains unclear.

Grants to extramural investigators are likely to be hit particularly hard, because they can generally be scaled down more quickly than internal costs such as salaries or long-term programme commitments. "You can't cut everything right away," says Jennifer Zeitzer, director of legislative relations for the Federation of American Societies for Experimental Biology in Bethesda, Maryland.

SEEKING CLARITY

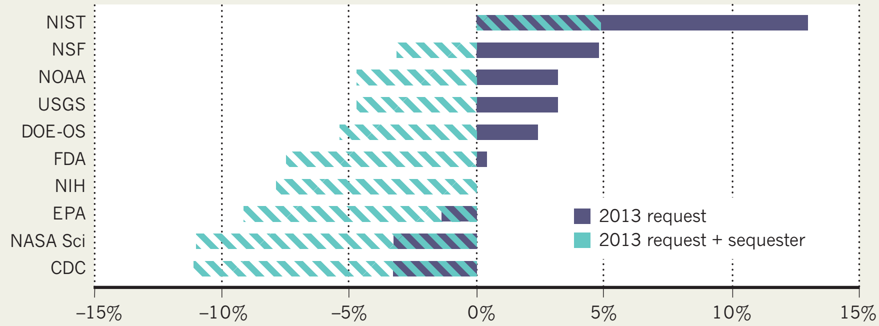
On 18 July, the Republican-led House passed 'transparency' legislation that would force President Barack Obama's administration to specify how the cut would be applied. If the Senate were to pass a parallel bill, the administration would be forced to provide the details within 30 days.

"Hopefully, with the transparency bill we'll get some answers on how to allocate those cuts," says Matt Hourihan, director of the research and development budget and policy programme at the American Association for the Advancement of Science in Washington DC.

Research advocates are particularly concerned that the sequester might be adjusted to prevent cuts to defence. If that happened, non-military programmes would be forced to bear more than twice the currently mooted cut. Such an extreme measure would threaten entire facilities and sideline thousands of research grants. "A lot of programmes wouldn't survive. They'd be vaporized," says Lubell.

WINNERS TO LOSERS

If a 2013 budget sequester enacts a predicted across-the-board cut of 8%, then most of the US science agencies expecting to see modest gains next year will instead experience deep cuts.



With the sequester lurking in the background, the latest move in the annual budget chess match came on 17 July, when the House subcommittee that oversees the budgets for the NIH and the Centers for Disease Control and Prevention in Atlanta, Georgia, released its version of a spending bill for fiscal year 2013 (see 'State of play'). For the NIH, which has a US$30-billion annual budget and is the biggest non-military research funder in the United States, the House bill offers little comfort. Overall, the legislation keeps the agency's funding at the same level as in 2012 — a net decrease once inflation is taken into account. This matches Obama's 2013 budget request to Congress in February (see *Nature* **482,** 283–285; 2012), but contrasts with the Senate version of the spending bill, which boosts the agency's funding by about $100 million.

With a trillion-dollar federal deficit at the forefront of legislators' minds, a clamp on spending is not as surprising as the unusually hands-on way in which the NIH portion of the House bill is written. The bill specifically requires NIH director Francis Collins to spend 90% of the agency's budget on extramural activities, 10% on intramural activities and at least 55% on basic science, and instructs the agency to maintain at least 16,670 training research awards (a category intended for doctoral students and postdoctoral fellows). That is roughly how many awards the agency currently makes, but fixing the number in the bill reduces the NIH's flexibility. The bill also prohibits the NIH from funding research that compares the relative effects of treatments on patients, part of a broader move by the House to thwart Obama's health-care reforms. Finally, the bill restricts travel by Collins and his staff until the NIH implements a pilot study on the medical-insurance coverage of patients in clinical trials that was requested by Congress last year.

Spending bills introduced in the House and Senate from April offer a brighter year for some other agencies: the rises proposed for the US National Science Foundation and the US National Institute of Standards and Technology at least approach those called for in the president's request (see 'State of play'). But the Department of Energy's Office of Science faces a stiff cut in the House bill as do the Food and Drug Administration and the Environmental Protection Agency, owing to political differences between House Republicans and the Obama administration. The reverse is true for the NASA science budget, with the House aiming to resist an administration cut that would jeopardize a sample-return mission to Mars.

In theory, the 2013 budget must be enacted by 30 September, the end of the 2012 fiscal year. That would require the House and Senate to agree on a bill and Obama to sign it. In practice, however, Congress is not expected to settle the budget until after the general election on 6 November. But the outcome will be just the tip of the fiscal iceberg if Congress cannot avert the sequester.

"Conventional wisdom at this point is that the sequester won't happen," says David Moore, senior director of government relations at the American Association of Medical Colleges in Washington DC. "But when you press legislators on what's going to happen to avert this, there's no consensus." ■

STATE OF PLAY

A comparison of available 2013 US budget figures for selected science agencies (in US$ millions). Some House and Senate numbers are yet to be voted on.

Agency	2012 enacted	2013 request	2013 House	2013 Senate
National Institutes of Health*	30,623	30,623	30,623	30,731
Centers for Disease Control and Prevention†	6,883	6,660	6,068	7,234
Food and Drug Administration	2,506	2,517	2,481	2,530
National Science Foundation	7,032	7,372	7,333	7,273
NASA (science)	5,074	4,911	5,095	5,021
Department of Energy Office of Science	4,874	4,992	4,801	4,909
National Institute of Standards and Technology	761	860	830	826
Environmental Protection Agency	8,450	8,344	7,055	Not tabled
National Oceanic and Atmospheric Administration	4,894	5,054	4,962	3,419‡
US Geological Survey	1,068	1,102	967	Not tabled

Sources: Office of Management and Budget; House and Senate appropriation committees; Centers for Disease Control and Prevention. *Includes only Department of Health and Human Services allocation. †Includes prevention and public-health fund. ‡Includes moving of satellite budget to NASA.

The Aquarius Reef Base enables ocean scientists to spend up to two weeks at a time living on the sea bed, conducting research on the coral reef nearby.

OCEANOGRAPHY

End of an age for Aquarius

Iconic underwater research base set to close its airlocks for the last time this year.

BY MARK SCHROPE IN KEY LARGO, FLORIDA

Floating 17 metres below the surface of the Atlantic, oceanographer and environmentalist Sylvia Earle is gazing affectionately at a loggerhead turtle dubbed Little Joe. Earle often campaigns to save endangered marine creatures such as loggerheads, but now she wants to protect a steel cylinder that looms not far behind her: the Aquarius Reef Base, which for the past week has been Earle's home on the sea bed just off Florida's Key Largo.

Two weeks living in Aquarius allows scientists to stay close to a coral reef, and to spend as much time underwater as they would in a year of sequential dives. Marine biologists say that no other facility gives them a chance to study reef ecosystems so intimately for extended periods. "You become a resident, not a visitor," says Earle. But to the dismay of many, Aquarius looks set to close its airlocks for good this year — a victim of tight budgets and waning political support.

The base has been operating in Florida since 1992, and at just over 13 metres long and 5 metres wide it comfortably accommodates a crew of six. Owned and largely funded by the US National Oceanic and Atmospheric Administration (NOAA) as part of its National Undersea Research Program (NURP), Aquarius is run by the University of North Carolina Wilmington. Over the past decade, the facility's budget has ranged from about US$800,000 to $3 million.

But the Aquarius team got a shock in US President Barack Obama's budget request for 2013 this February. NOAA had recommended that Congress eliminate funding for Aquarius and terminate NURP. "Never once did they mention that this was coming down," says Thomas Potts, Aquarius's director.

➲ NATURE.COM
Watch a video of Mark Schrope's visit to Aquarius:
go.nature.com/dqhg17

To draw attention to the lab's plight, Potts and Mark Patterson, a long-time Aquarius user from the Virginia Institute of Marine Science in Gloucester Point, planned a high-profile final mission. They hoped to win back government support or attract private donations to a new Aquarius Foundation that would support the facility. Patterson recruited Earle, a former NOAA chief scientist who consulted on Aquarius's location two decades ago and is now an explorer-in-residence for the National Geographic Society; oceanographer Dale Stokes from the Scripps Institution of Oceanography in La Jolla, California; and underwater filmmaker D. J. Roller to join the mission. Countless reporters and members of the support crew also visited the site on each day of the 15–21 July expedition.

Earle and Patterson say that the lab makes a unique contribution to ocean science because it is the only place where researchers can stay on the bottom for long stretches — saturation diving — and reach surrounding undersea study sites in mere minutes. They and other supporters point to a long list of research accomplishments, including assessing the

role of sponges in filtering water around reefs (see *Nature* **457,** 141–143; 2009); the discovery of huge waves of cool water from deeper areas offshore that wash over reefs (J. J. Leichter *et al. J. Phys. Oceanogr.* **35,** 1945–1962; 2005); and the development and testing of instruments that enable such work. Supporters also say that the lab is invaluable for ocean outreach. And since 2001, NASA has been renting Aquarius as a training proxy for space operations, because working in water offers one of Earth's closest analogues to working in microgravity.

But Andrew Shepard, who directed Aquarius from 2004 to 2009, was not surprised by the cut. "It's the age-old battle of extramural versus intramural costs," he says. In recent years, NOAA's Ocean Exploration and Research programme, which supports NURP, has struggled to adequately fund core priorities such as the ship *Okeanos Explorer*, and Shepard says that extramural programmes such as Aquarius are typically a lower budget priority. He also acknowledges that "not everybody thinks that Aquarius is indispensable to the coral-reef-science community. Some people say it's in the wrong place, that it might be useful if you could move it around."

"I think the big questions are, should it keep going, and if so, for how much longer, and what should it look like?" he adds.

Aquarius has faced major budget crises before, including complete funding elimination in past presidential budget requests. But congressional supporters have always restored funding. Now, with the current drive to rein in government spending, "that's taboo", says Shepard.

NOAA declined *Nature*'s requests for an interview, but in an e-mailed statement, spokesman David Miller said that Aquarius has been a "vital part" of fulfilling the agency's core missions. "Unfortunately, our budget environment is very, very challenging and we are unable to do all that we would like," he adds. Although the president's 2013 budget request calls for an increase of almost $160 million for NOAA, much of that is designated for weather-satellite programmes.

Last week's mission did generate substantial attention, including visits from several Florida legislators. But barring the arrival of a last-minute saviour, the Aquarius team will disband in December.

Joe Deppen, a habitat technician and diver for Aquarius for the past three years, is already planning to train at a commercial diving school. "I wish people would realize what an asset Aquarius is," he says, pausing to consider his words while rocking on a support boat in choppy seas just above the base. "Everyone that comes down realizes what an amazing place this is." ■

GENOMICS

Contest to sequence centenarians kicks off

First entrant pins hopes on semiconductor technology.

BY MONYA BAKER

The first competitor has swaggered up to the starting line for a contest that aims to push the limits of genome-sequencing technology. The X Prize Foundation of Playa Vista, California, is offering a US$10-million prize to the first team to accurately sequence the genomes of 100 people aged 100 or older, for $1,000 or less apiece and within 30 days. Ion Torrent, part of Life Technologies of Carlsbad, California, believes that its semiconductor-based technology gives it a shot, and on 23 July it announced that it will compete.

The Archon Genomics X Prize competition, to be held in September 2013, is intended to spur technology, boost accuracy and drive down costs — currently $3,000–5,000 per genome. Peter Diamandis, the X Prize Foundation's chief executive, says that the contest will help to establish a standard for a "medical grade" genome, with the high accuracy needed to diagnose or treat a patient. Eventually, says Michael Snyder, a geneticist at Stanford University in California, "I do see a world where you'll have a funny-looking mole and they'll pull that off and want to sequence its genome."

The contest will also test the claims that sequencing companies have made. "I call it the truth serum for the genome-sequencing field," says genome maven Craig Venter, co-chairman of the competition.

Entries can be made until May next year, and organizers hope that the competition will be keener than in past events. An earlier incarnation of the competition, launched in 2006, drew several registrants, but none came close to winning.

This time, the X prize Foundation has relaxed the time frame, allowing competitors 30 days — rather than the 10 specified by the 2006 contest — and focused on centenarians, who might carry gene variants promoting longevity. The winning team will be the first to sequence all 100 genomes to 98% completion, with less than one error per million base pairs, and to determine which variants appear on which of the paired chromosomes. In case there is no outright winner, lesser 'best in class' awards are on offer. "It would surprise me if one team gets all of the criteria," says Kevin Davies, editor-in-chief of *Bio-IT World* and author of the book *The $1000 Genome* (Free, 2010).

Ion Torrent believes that its edge will come from its technology, which measures a tiny change in pH each time a specific base is added to a growing DNA strand. Most sequencing technologies rely on light emitted as bases are incorporated, which requires higher built-in costs for optical equipment.

Davies regards Oxford Nanopore Technologies in Oxford, UK, as a potential rival. Earlier this year, the company said that it would soon be able to sequence a human genome in 15 minutes. Its technology threads a DNA strand through a nanometre-scale hole and senses each base as it passes through. But the company would not say whether it intends to enter the competition. Spokespeople at the sequencing-services companies BGI in Shenzhen, China, and Complete Genomics in Mountain View, California, said that the firms had not yet decided whether to enter.

A poster used to recruit DNA donors.

X PRIZE FOUNDATION

Clifford Reid, chief executive of Complete Genomics, worries that it will be difficult for the judges to assess the accuracy of the newly sequenced genomes. "The technologies participating in the competition are the only technologies for judging the competition," he says, adding that he is hopeful that contest organizers can come up with "a clever solution that makes everyone happy".

Although the contest will reward technological prowess, Venter says the key challenge is not amassing sequence, but understanding what it means for biology and medicine. "The trivial part of the equation to solve is the sequencing technology," he says. "It's necessary, but not sufficient." ■

POLITICS

US drug agency spied on scientists

Food and Drug Administration monitored five employees, defying promises about whistle-blower protection.

BY MEREDITH WADMAN

Until last week, the US Food and Drug Administration (FDA) had been remaking its image as a transparent organization that was supportive of its scientists, even when they spoke out against its decisions.

Now the agency is on the defensive, after the exposure of a clandestine computer-surveillance operation that tracked every keystroke made by five dissident FDA scientists whom it suspected of leaking confidential internal data to the press. The revelation may damage employees' trust in the FDA, and erode their willingness to challenge the decisions of their bosses, say expert observers. "The mere act of monitoring e-mails can chill scientific discourse at the agency and leave scientists more vulnerable to retaliation," says Michael Halpern, the integrity programme manager at the Union of Concerned Scientists (UCS), an advocacy group based in Cambridge, Massachusetts.

UCS surveys of more than 900 FDA scientists had shown that the proportion who feared retaliation for openly expressing concerns about the agency's work fell from 36% to 26% between 2006 and 2011. In the same period, the proportion who said that their supervisor "stands behind scientists" who put forth controversial views climbed from 38% to 61%. And when the agency issued a scientific-integrity policy in February, "supporting whistleblower protections" was on a list of key principles. Another principle read, "Allowing FDA staff to communicate their personal scientific or policy views to the public, even when those views differ from official Agency opinions."

Critics say that the surveillance campaign strikes at the very heart of those principles. In addition to monitoring keystrokes, the FDA used software to capture all data stored on the computers and on USB sticks, and all e-mails sent and received on the computers, whether using personal or government accounts. The software also took screenshots at five-second intervals. Writing to FDA commissioner Margaret Hamburg last week, Senator Charles Grassley (Republican, Iowa), who is investigating the surveillance, alleged that the operation had been "explicitly authorized, in writing" by the FDA's head lawyer. The agency gathered more than 80,000 pages of information during the operation, says *The New York Times*, which first disclosed the extent of the surveillance on 14 July.

The scientists, who worked in the FDA's Center for Devices and Radiological Health in Silver Spring, Maryland, first began to speak out in 2008, telling the US Congress that the agency's process for reviewing medical devices was "corrupted".

An article about a breast-cancer imaging device in *The New York Times* in January 2009 was followed by one in March 2010 that quoted an internal review by one of the scientists. The review cautioned against the FDA's potential approval of a colon-cancer screening device that the scientist believed delivered dangerous levels of radiation (see 'Under surveillance').

The FDA says that the disclosures to the newspaper were illegal under the Federal Food, Drug, and Cosmetic Act, which prohibits the agency from publicizing data and information submitted by a drug- or device-maker before marketing approval has been granted — even the existence of an application cannot be disclosed. GE Healthcare of Little Chalfont, UK, had applied to have a device approved for routine colon-cancer screening among people without symptoms — a huge and lucrative market. After the 2010 article appeared, the company wrote to the FDA asking it to investigate how, as *The New York Times* had reported, "scores of internal agency documents" concerning its application had been leaked to the newspaper.

In a letter to Grassley on 13 July, the FDA said that it began monitoring the scientists' government-owned computers in April 2010. Erica Jefferson, a spokeswoman for the agency, said that the monitoring "was only intended to identify the source of the unauthorized disclosures, if possible, and to identify any further unauthorized disclosures". By the end of 2011, four of the five scientists had been fired or had not had their contracts renewed. Ewa Czerska, who had been at the agency for 23 years, was dismissed "for unauthorized disclosure of confidential information", the agency wrote in its letter. It did not describe its reasons for terminating the others' employment.

The scientists sued the FDA in January, claiming that the agency had violated their rights to free speech and association, their right to petition Congress and their right to be protected from unreasonable search and seizure. In a revised lawsuit filed last week, they also allege that the monitoring actually began in 2009, and that "the FDA intercepted private e-mails that were composed during non-work hours, from home, on personal networks and non-government computers". Jefferson says that the monitoring was limited to the five employees' government-owned computers; Hamburg declined to be interviewed.

The captured documents include personal communications, among them lawyer–client exchanges, as well as letters to Congress and the government's Office of Special Counsel, which investigates whistle-blower complaints and is meant to protect whistle-blowers from retaliation. "This case is different from any we have seen in the past because of the sweeping and pervasive nature of the surveillance conducted, and because the scientists were using laptop computers both at home and at work for a variety of personal and private purposes," says Alan Butler, a privacy-law expert at the Electronic Privacy Information Center in Washington DC.

"These employees are properly going to members of Congress or the Office of Special Counsel and they are being retaliated against, presumably as a result," says Mark Zaid, a lawyer in Washington DC, who specializes in defending whistle-blowers. "It sends a chilling message." ■ SEE EDITORIAL P.405

UNDER SURVEILLANCE

The US Food and Drug Administration (FDA) monitored employees after they voiced concerns about imaging devices.

28 MARCH 2010 *The New York Times* says that FDA managers suppressed scientists' concerns about radiation risks from routine colon-cancer screening.

16 APRIL 2010 GE Healthcare alleges that confidential proprietary information had been leaked.

22 APRIL 2010 The FDA starts to put spyware on scientists' computers.

JULY 2010–OCTOBER 2011 Four monitored scientists lose jobs.

25 JANUARY 2012 Scientists sue the FDA for violating their civil rights.

14 JULY 2012 *The New York Times* reports that the FDA has amassed more than 80,000 documents during the surveillance campaign.

16 JULY 2012 Senator Charles Grassley calls for investigation.

PHYSICS

US fusion in budget vice

Domestic facilities struggle for survival as funding is directed to international reactor.

BY ERIC HAND

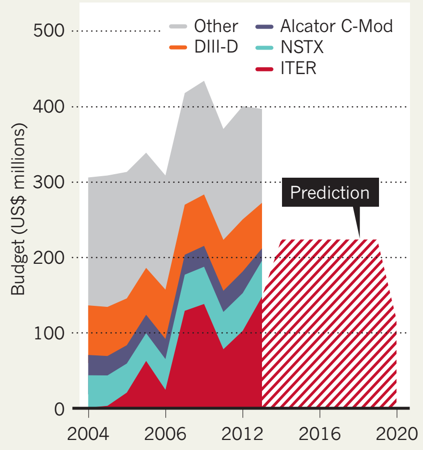
For years, US researchers have been steadfast in their support of ITER, the world's largest fusion-energy experiment, which is under construction near Cadarache, France. But with funding commitments to ITER now putting the squeeze on three existing facilities in the United States, enthusiasm for the international project is becoming as difficult to sustain as a fusion reaction.

"I think we should ask whether this is the right path," Earl Marmar, head of the Alcator C-Mod fusion experiment run by the Massachusetts Institute of Technology in Cambridge, told colleagues on 18 July. The venue was a meeting of a US Department of Energy (DOE) group tasked with setting priorities for the non-ITER portion of the US fusion programme. At the meeting, in Bethesda, Maryland, Marmar pointed out that when US fusion researchers signed on to ITER in 2003, the project's total construction cost was projected to be about US$5 billion, of which the United States would provide 9% over ten years. Now, the construction costs are projected to be roughly four times as much. Furthermore, the funds to support ITER were not supposed to be siphoned from existing facilities — yet if the total budget for US fusion science remains flat, as is expected, that is precisely what will happen (see 'Death by ITER').

Marmar's facility houses one of three US tokamaks — doughnut-shaped vessels in which physicists magnetically confine hydrogen nuclei in a plasma and heat them until they fuse and liberate energy. Alcator received $29 million in federal funding this year. But as ITER payments increase, US President Barack Obama's 2013 budget proposal for the DOE would chop Alcator's allocation back to $16 million, shutting down operations and forcing the experiment to lay off more than half of its 120 staff members.

DEATH BY ITER

If the budget for US fusion science remains flat, rising commitments to the international reactor ITER mean that the three domestic reactors would be squeezed to the point of extinction.



SOURCE: D. BRUNNER/MIT

Stephen Dean, president of Fusion Power Associates, an advocacy group in Gaithersburg, Maryland, says that DOE officials have little choice but to cut Alcator, the smallest of the three US experiments, to afford an overall US ITER commitment that has grown to about $2.2 billion. "Why can't we get by with two?" asks Dean. "It's not an insubstantial argument."

➲ NATURE.COM
For more about fusion projects in the United States, visit:
go.nature.com/jspfs7

Yet leaders at the three experiments insist that all provide unique science (see 'Tokamaks under pressure'). Marmar says that Alcator, for example, can operate at extremely high magnetic-field strengths that mimic those planned at ITER. And whereas most tokamaks have inner walls made of graphite, Alcator researchers have pioneered the use of tungsten — a more durable material that ITER is planning to adopt. Current experiments at Alcator also explore the use of special radio-wave antennas to heat the plasma in ways that reduce erosion of the walls.

Stewart Prager, director of the Princeton Plasma Physics Laboratory in New Jersey, has a different argument for keeping all three experimental facilities running. If the United States is to spend all this money on ITER, he says, then it must maintain domestic plasma-science expertise that can take advantage of what is learned there. "Otherwise, the results from ITER will only benefit the rest of the world."

It seems that some members of the US Congress are listening. On 6 June, the House of Representatives voted to boost ITER funding and to support the domestic programme at almost 2012 levels. The Senate's version of the bill, which has not yet been voted on, currently agrees with the cuts in the Obama administration's budget request — but directs the DOE to explore the impact of simply withdrawing from ITER.

US fusion researchers do not want that — yet. But if the 2014 budget looks at all like the 2013 one, Dean predicts, the knives will be out for ITER. "They're not trying to kill ITER just yet," he says. "If this happens again in 2014, I'm not so sure."■

TOKAMAKS UNDER PRESSURE

US President Barack Obama's 2013 budget proposal for the Department of Energy would cut funding for the three US tokamak fusion reactors by US$24 million and put one, Alcator C-Mod, on a path to cancellation.

Name	Location	Year completed	2012 funding	Plasma volume	Peak magnetic field	Peak duration of plasma heating	Unique science
DIII-D	San Diego, California	1986	$69 million	20 cubic metres	2.2 tesla	10 seconds	Because DIII-D has a similar shape to the international ITER reactor, it is a test bed for working out design kinks in the larger project.
Alcator C-Mod	Cambridge, Massachusetts	1993	$29 million	1 cubic metre	8 tesla	4 seconds	Alcator pioneered tokamak walls made of tungsten, an approach now being used at ITER.
National Spherical Torus Experiment (NSTX)	Princeton, New Jersey	1999	$50 million	11 cubic metres	1 tesla	5 seconds	The nearly spherical plasma could allow fusion in lower magnetic fields than required by other reactors — and with lower magnet costs.

Source: DIII-D/Alcator C-Mod/NSTX

THE TIME MACHINE

Dating features on the Moon and Mars is guesswork. Scott Anderson is building a tool to change that.

BY ERIC HAND

The bits of rock on Scott Anderson's shelf are not much to look at, but they have stories to tell. In a plastic case is a greenish-grey rock, a 4.5-billion-year-old piece of the asteroid Vesta. Next to it rests a dark sliver of 2.8-billion-year-old lava from the Moon. Anderson, a planetary scientist at the Southwest Research Institute in Boulder, Colorado, picks up his favourite, a 1-gram slice of rock that cost him US$800. The flake came from Zagami, an 18-kilogram meteorite named after the Nigerian village where it was found in 1962. It is one of the rarest and most sought-after types of meteorite — a piece of Mars that was blasted into space by an asteroid impact and eventually landed on Earth. "Knowing what it is makes me excited to see it every time," Anderson says.

What Anderson wants from these far-flung fragments of the Solar System is elementary: their ages. Coaxing out that information is far more difficult. Zigzagging across his laboratory is a web of laser beams that feed into a mass spectrometer — all part of a geochronometer that Anderson is building. Like other rock-dating systems, this one computes an age from the radioactive decay of certain isotopes in a sample. What sets Anderson's system apart is his goal to shrink the whole operation down to something that would fit on a desktop. Then, rather than waiting for planetary fragments to fall to Earth, he wants to send his device to the planets.

Over the past few decades, planetary scientists have mapped the Solar System in ever more staggering detail. Cameras orbiting the Moon and Mars can zoom in on objects as small as dinner plates, and radars can penetrate several metres below the surface. But when it comes to the fourth dimension — time — they are as blind as ever. Scientists have hard dates for only nine places in the Solar System, all on the Moon: six Apollo sites and three Soviet Luna sites, from which samples were returned robotically. When did water flow on Mars? When did the Moon's volcanoes last erupt? Without dates, planetary scientists can only make educated guesses about some of their most pressing questions.

A portable, *in situ* chronometer such as Anderson's could revolutionize how researchers study the Moon, Mars or other rocky bodies. The costs of big planetary missions are skyrocketing; the $2.5-billion Mars Science Laboratory that is scheduled to land on 6 August is one of the most expensive Mars missions ever. But Anderson's tool could reduce future costs, in particular by avoiding the need for budget-busting missions to retrieve samples from other planets and haul them back to Earth. And the device could even find a wide audience on Earth, among geologists who could use it to map the ages of rocks in the field, rather than delivering samples to a lab and waiting months for the results.

Scott Anderson plans to finish the prototype for his portable geochronometer later this year.

MATTER OF SCALE

But first, Anderson has to transform the finicky set-up that sprawls across his lab into one that could fly in space. Other groups are trying to develop portable geochronometers, but Anderson's design has some advantages, and he is closer to completing a working prototype. At present, the half-built apparatus sits in the corner of his office: 160 kilograms of gleaming steel and aluminium, roughly the size of a two-drawer filing cabinet. He hopes to finish it later this year, and then he will bolt it into the back of a van and take it on a road trip. "We've been talking about how we could drive this to NASA headquarters and test this in the parking lot," says Anderson. At 44 years old, he is tall and boyishly earnest, but savvy enough to understand good public relations. He wants to persuade NASA officials to pay to build an ultra-lightweight geochronometer and then send it on a rover to the Moon or Mars.

Anderson will have to show not only that his chronometer is fast and light, but also that his dates make sense. Radiometric dates are some of the trickiest, most delicate and most disputed measurements on Earth. Anderson wants to transform what has been a laborious process of chemical extraction and analysis into a laser-based system, automate it and shrink it into a robot small and reliable enough to send

➲ NATURE.COM
For more on Mars and the Curiosity mission, see: go.nature.com/fknipi

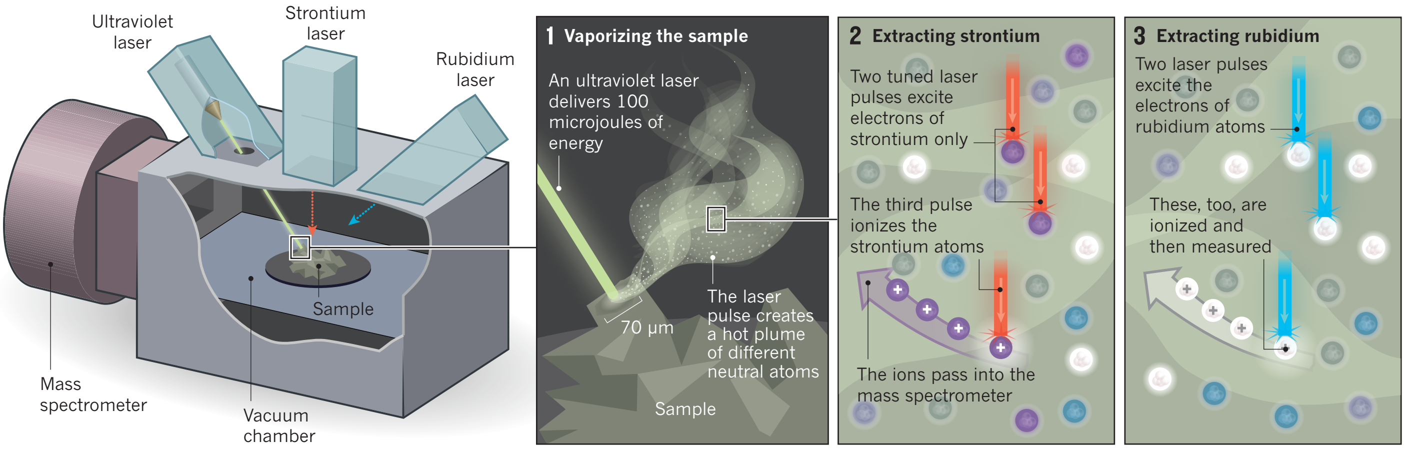
PHOTO BY MATT NAGER

SPEED DATING

Researchers at the Southwest Research Institute in Boulder, Colorado, are building a device that can quickly determine the age of rocks by measuring the decay of radioactive rubidium-87 to strontium-87. The tool is designed to be small and lightweight enough to fit on a mission to another planet.

The device uses three lasers to isolate and then ionize atoms of rubidium and strontium. Those ions are sent to a mass spectrometer, which measures their abundances.

The process requires a coordinated sequence of laser blasts, separated by just microseconds. The machine repeats this sequence thousands of times across the surface of a sample. The relative amounts of the strontium and rubidium isotopes reveal the age of the rock.



to another planet. "We're extremely sceptical of these things working," says Lars Borg, a chemist at the Lawrence Livermore National Laboratory in Livermore, California, whose three-person lab usually produces just two dates a year. "We really struggle to get these ages ourselves."

But this spring, Anderson used the full-sized version of his system to date a 1.7-billion-year-old piece of Boulder Creek granite that he chipped out of the foothills near his lab. Anderson's system computed an age of 2.05 billion years ± 130 million years — not great in terms of accuracy, but at least a proof of principle[1]. Next up will be Zagami — a precious rock sample that he does not want to put in the machine until it is ready.

Some researchers are now paying attention. "I would not have thought that they could progress this quickly," says Hap McSween, a planetary scientist at the University of Tennessee in Knoxville. "They're convincing me that there really is something to this."

QUESTION OF TIME

Anderson has wondered about the ages of rocks since he was a boy, when he would often tag along on field trips with his father, a sedimentologist at Temple University in Philadelphia, Pennsylvania. He learned about the principle of superposition — that younger layers are deposited on top of older ones — and how fossils can connect layers on different continents to a single epoch in the distant past. But the work was slow. "We spent hours staring at the same two feet of stone, getting sunburned and the bugs eating at you," he says.

As an undergraduate at Brown University in Providence, Rhode Island, Anderson discovered that he could avoid the insects by doing geology on other planets — where dates were even harder to come by. Planetary scientists had no fossils to work with, but on the Moon and Mars they had something else: thousands of craters left by large asteroids. In 1965, William Hartmann, a researcher at the Planetary Science Institute in Tucson, Arizona, developed a simple chronometer relying on the idea that surfaces marred by many craters should be older than ones with fewer blemishes. To date a surface, Hartmann used estimates of the rate of impacts over time, which he based on data collected on Earth[2].

This approach improved after the first Apollo rocks had been dated and the crater-count method was calibrated[3]. But even now, it yields dates with significant uncertainties — between 10% and 40%, Hartmann says. That is mostly because no lunar rock samples have been retrieved from surfaces between 1 billion and 3 billion years old. Scientists are eager to fill in that gap.

If chronology on the Moon is still uncertain, then Mars is a mess. The crater-count method does not work as well there, mainly because the wind, water and frost that sculpt the surface also erase craters. Translating the Moon's crater chronometer to Mars is a delicate business, says Barbara Cohen, a planetary scientist at Marshall Space Flight Center in Huntsville, Alabama, who is developing a rival portable chronometer. "On Mars, and on every other planet, all we're doing is extrapolating, for better or for worse, with a fudge factor." That 'fudge factor' could be erased with a few choice dates.

But those dates could do much more than simply calibrate the crater chronometer. With a portable system, researchers could decipher how long volcanism lasted on Mars and when it stopped. They could find out when the planet's warm, wet and possibly habitable environment gave way to the cold desert it has been for several billion years. "If any evidence is found for life, we sure as heck will want to know when it was there," says McSween.

These questions are some of the reasons that a generation of scientists have sought a mission to retrieve rock samples from Mars. In 2011, a sample-return mission was ranked the top mission priority for planetary science in the US National Research Council's decadal survey. But NASA's budget-minders baulked at the price tag for the project, which would have required three separate missions and a sample-handling facility — costing more than $10 billion in total over a decade. Anderson's device would cut out the return trip and do the dating on site. A Mars rover of any size is not cheap, but a medium-sized rover might be possible for about $1 billion. At that price, multiple dating missions could be sent to multiple locations. Doing the science *in situ* also sidesteps the costs of building a quarantined facility on Earth to handle the samples.

Defenders of the sample-return approach argue that there are many reasons for such a mission, beyond the mere dating of rocks. Searching for life in the samples would be the top priority, and that would be easier to do in a lab on Earth. But if Anderson and his competitors can demonstrate the viability of their portable geochronometers, support for a more complex and costly sample-return mission could diminish quickly. "I'd rather have five of those ages from five places on Mars than one sample return," says Hartmann.

Even before reaching the red planet, Anderson's device could win fans here on Earth. Geologists typically spend weeks or months out in the field and then haul sacks of rocks back from remote places for extensive analysis. Sometimes much of the effort is for naught — the samples may be unsuitable for dating, or the researchers may have

picked up rocks that were older or younger than the period they wanted to study. A portable geochronometer that could produce a date within hours could solve those problems.

At the moment, some technical issues stand in the way. Anderson has spent the summer waiting for the delivery of a $200,000 laser that he needs to complete the device. In the meantime, he's had to jury-rig his lab so that several of his older lasers, cooled by a roomful of refrigerators and water pumps, shine into a vacuum chamber affixed to the brand-new, $700,000 mass spectrometer that will form the bulk of the portable prototype. One of the lasers, nicknamed Jill, is newly rehabilitated after a problem that announced itself with an acrid, burning smell. "Our best guess is something crawled in and committed suicide," says Keith Nowicki, a laser physicist in the lab.

In principle, Anderson's radiometric dating technique is similar to any other. It is based on the radioactive decay of one isotope into daughter isotopes according to a precise clock, a half-life, governed by nuclear physics. Anderson's method relies on rubidium-87, which decays to strontium-87 with a half-life of 48.8 billion years. This method, like any radiometric technique, typically requires monumental efforts. Researchers must first crush the rock and separate its minerals, often by hand. The minerals must then be dissolved in a strong acid, which goes through cation-exchange columns to extract the radioisotopes. These are dried and their abundance measured in a mass spectrometer. The steps can take months to complete. "The process is a pain in the neck," says Anderson.

LASER POWER

Anderson's method avoids some of these hassles by using tunable lasers to liberate and sort the isotopes all at once. During a visit, Nowicki and Anderson demonstrate the system on a piece of Boulder Creek granite. Even reflected light from the ultraviolet lasers is strong enough to blind, so the researchers first put on thick, $600 protective goggles that dim the room and colour it a sickly ochre.

Nowicki turns the lasers onto a wafer-thin slice of rock. Instantly, values for the abundances of rubidium and strontium appear as curves on a computer screen. Anderson is constantly tweaking the protocol for determining a date, but it always involves three basic steps (see 'Speed dating'). First, a blast of laser light vaporizes a smidgeon of the rock sample, creating a cloud of neutral atoms and a pit 70 micrometres around. Next, another, precisely tuned laser fires two shots, nanoseconds apart, to excite only the electrons in the strontium atoms in the cloud. A third shot rips those electrons away, turning the atoms into ions that are then whisked into the mass spectrometer. A microsecond later, three finely tuned shots ionize the rubidium atoms (which are still lingering in the cloud), and these are sucked into the mass spectrometer and measured. The process is repeated 20 times a second — and 3,000 times in the same place — and then the laser is pointed at a new spot on the sample's face. To date an entire sample, Anderson usually measures several hundred spots, which takes about a day and a half.

Other researchers, such as Cohen and John Eiler, a geochemist at the California Institute of Technology in Pasadena, are trying to develop *in situ* geochronometers that use potassium–argon dating, in which potassium-40 decays to argon-40 with a half-life of 1.3 billion years. Argon, a noble gas, tends to remain trapped in the crystal matrix of minerals. Potassium and argon are more abundant in common minerals than rubidium and strontium, which makes them easier to measure. But the potassium–argon system does not work as well for rocks that have been disturbed by high pressures and temperatures, which can cause argon to leak out and make the rocks seem younger than they are. And samples from Mars could have the opposite problem: argon-40 in the planet's atmosphere and mantle might have seeped into rocks, artificially inflating their ages.

Because each system has unique advantages and disadvantages, it may be best to put a couple of portable geochronometers on a rover, so that the results can be checked against each other. So Anderson has been busy forging alliances with Cohen and other former competitors to develop a viable mission proposal for a chronometer-laden rover.

They just might have a shot. In February, under intense budget pressure, NASA threw out its $10-billion, long-term Mars plan that would have begun a sample-return mission at the end of the decade. The new plan leaves only about $800 million for a Mars mission in 2018 or 2020, just enough for an orbiter, lander or, perhaps, an inexpensive rover.

MAJOR-LEAGUE PITCH

One day in June, Anderson flies to Houston, Texas, where NASA officials are holding a conference to solicit ideas for the new mission. Dozens of concept studies are vying for the attention of officials: robots that climb rocks, rovers that hop and autonomous skiffs that would explore Mars as they are whisked along by the wind.

Anderson gets ten minutes to present his team's concept. His chronometer would be mounted on an enhanced version of the Mars Exploration Rovers — Spirit and Opportunity — that landed in 2004. The new rover would have a life-detection experiment on board and might have room for a small sample cache, to preserve the chance of a future sample-return mission. Cohen's potassium–argon system would also squeeze aboard. By the time Anderson gets round to explaining his part — the rubidium–strontium geochronometer — he has three minutes to talk about the thing he has been working on for eight years.

> "IF ANY EVIDENCE IS FOUND FOR LIFE, WE SURE AS HECK WILL WANT TO KNOW WHEN IT WAS THERE."

He tells the audience that his rover would do important science and lay the groundwork for a sample return, without an absurdly high price tag. "More science, less commitment," he says.

Later, Anderson says that he has no idea how his concept was received by the NASA officials. "They're holding their cards pretty close to the vest." He knows that winning the flight opportunity is a long shot, so he is thinking about fallbacks. In 2015, NASA expects to solicit proposals for low-cost planetary missions, and Anderson plans to pitch sending a geochronometer to the Moon. But the bigger target is never far from his thoughts.

Anderson recently bought a 28-centimetre telescope — pretty big for an amateur — and installed it in his backyard in Boulder, which sits at 2,200 metres and has a clean view of the sky through the thin Rocky Mountain air. He has spent many an evening staring at the red planet — and imagining his timepiece at work on its surface. "I want to get it to Mars," he says. "I want to see it there." ■

Eric Hand *covers physical sciences for* Nature *in Washington DC.*

1. Anderson, F. S. *et al.* Presentation at the NASA Concepts and Approaches for Mars Exploration meeting, Houston, Texas; 2012. Available at http://go.nature.com/ahgamj.
2. Hartmann, W. K. *Icarus* **4,** 157–165 (1965).
3. Hartmann, W. K. *Icarus* **13,** 209–301 (1970).

COMMENT

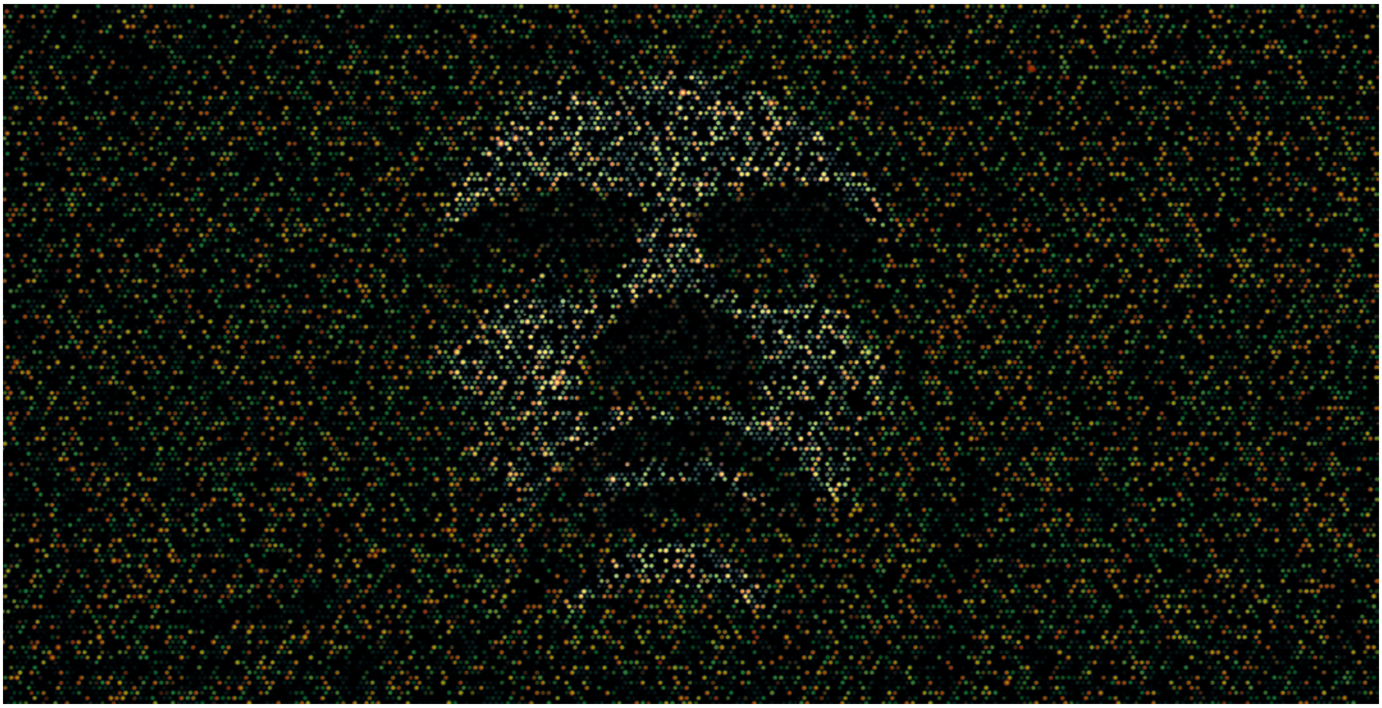
VISUALIZATION Lavish lab guide lays out steps to making images that wow **p.430**

ART Infrared analysis shows Albrecht Dürer worked less on pictures destined for sale **p.431**

PUBLISHING Ethics procedures should be required in papers, like other methods **p.432**

FUNDING Patients' priorities for schizophrenia research get UK support **p.432**

SHUTTERSTOCK/W. FERNANDES



Face up to false positives

Scientists and journals must work together to ensure that eye-catching artefacts are not trumpeted as genomic insights, says **Daniel MacArthur**.

When a study of the genomes of centenarians reported genetic variants strongly associated with exceptional longevity[1], it received widespread media and public interest. It also provoked an immediate sceptical response from other geneticists. That individual genetic variants should have such large effects on a complex human trait was totally unexpected. As it turned out, at least some of the results from this study were surprising simply because they were wrong. In a retraction published a year later[2], the authors admitted to "technical errors" and "an inadequate quality control protocol". The work was later republished in a different journal after heavy revision[3].

Few principles are more depressingly familiar to the veteran scientist: the more surprising a result seems to be, the less likely it is to be true. We cannot know whether, or why, this principle was overlooked in any specific study. However, more generally, in a world in which unexpected results can lead to high-impact publication, acclaim and headlines in *The New York Times*, it is easy to understand how there might be an overwhelming temptation to move from discovery to manuscript submission without performing the necessary data checks.

In fact, it has never been easier to generate high-impact false positives than in the genomic era, in which massive, complex biological data sets are cheap and widely available. To be clear, the majority of genome-scale experiments yield real results, many of which would be impossible to uncover through targeted hypothesis-driven studies. However, hunting for biological surprises without due caution can easily yield a rich crop of biases and experimental artefacts, and lead to high-impact papers built on nothing more than systematic experimental 'noise'.

Flawed papers cause harm beyond their authors: they trigger futile projects, stalling the careers of graduate students and postdocs, and they degrade the reputation of genomic research. To minimize the ▶

damage, researchers, reviewers and editors need to raise the standard of evidence required to establish a finding as fact.

Two processes conspire to delude ambitious genomicists. First, the sheer size of the genome means that highly unusual events occur by chance much more often than we would intuitively expect. The limited grasp of statistics that many biologists have and the irresistible appeal of biological findings that neatly fit the facts are a recipe for spurious findings.

Second, all high-throughput genomic technologies come with error modes and systematic biases that, to the unwary eye, can seem like interesting biology. As a result, researchers who are inexperienced with a technology — and some who should know better — can jump to the wrong conclusion.

Again, whether these factors play a part in any specific case is often impossible to know, but several high-profile controversies highlight the potential impact of chance and technical artefacts on genome-scale analyses. For instance, rare loss-of-function mutations in a gene called *SIAE* were reported to have a large effect on the risk of autoimmune diseases[4]. But a later, combined analysis of more than 60,000 samples[5] showed no evidence of an association, suggesting that the finding in the original publication was down to chance. Key results in the retracted genetic analysis of longevity mentioned earlier[1] turned out to be errors that arose as a result of combining data from multiple genotyping platforms. And a study published last year that reported widespread chemical modification of RNA molecules[6] was heavily criticized by experts, who argued that the majority of claimed modifications were, in fact, the product of known classes of experimental error[7–9].

Resolving such controversies after results have been published can take time. Even after a strong consensus has emerged among experts in the field that a particular result is spurious, it can take years for that view to reach the broader research community, let alone the public. That provides plenty of opportunity for damage to be done to budding careers and to public trust.

REPLICATION AND REVIEWING

How can the frequency with which technical errors are trumpeted as discoveries be minimized? First, researchers starting out in genomics must keep in mind that interesting outliers — that is, results that deviate significantly from the sample — will inevitably contain a plethora of experimental or analytical artefacts. Identifying these artefacts requires quality-control procedures that minimize the contribution of each to the final result. Finding different ways to make data visual (including simply plotting results across the genome) can be more helpful than many researchers appreciate. The human eye, suitably aided, can spot bugs and biases that are difficult or impossible to see in massive data files. Crucially, genomicists should try to replicate technology-driven findings by repeating the study in new samples and using experimental platforms that are not subject to the same error modes as the original technology.

Stringent quality control takes time, a scarce resource in the fast-paced world of genomics. But researchers should weigh the risk of being scooped against the embarrassment of public retraction.

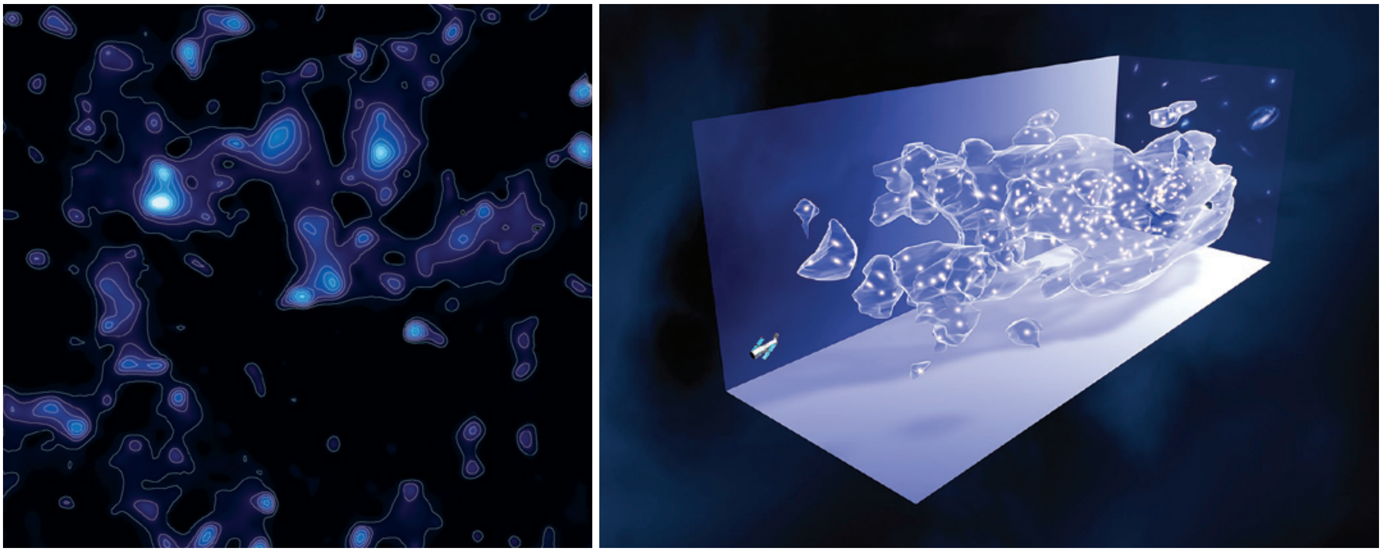
For 'paradigm-shifting' genomics papers, journal editors must recruit reviewers who have enough experience in the specific technologies involved to spot subtle artefacts. Often these will be junior researchers working in the trenches of quality control and manual data inspection. In addition to having the necessary experience, such reviewers often have more time for careful analysis than their supervisors.

Finally, the genomics community must take responsibility for establishing standards for the generation, quality control and statistical analysis of high-throughput data generated using new genomic technologies (a model that has generally worked well, for instance, in genome-wide association studies) and for responding rapidly to published errors. Traditionally, scientists wrote politely outraged letters to journals. Many now voice their concerns in online media, a more rapid and open way to ensure that the public view of a finding is tempered with appropriate caution. Such informal avenues for rapid post-publication discourse should be encouraged.

Nothing can completely prevent the publication of incorrect results. It is the nature of cutting-edge science that even careful researchers are occasionally fooled. We should neither deceive ourselves that perfect science is possible, nor focus so heavily on reducing error that we are afraid to innovate. However, if we work together to define, apply and enforce clear standards for genomic analysis, we can ensure that most of the unanticipated results are surprising because they reveal unexpected biology, rather than because they are wrong. ■

Daniel MacArthur *is based at the Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.*
e-mail: macarthur@atgu.mgh.harvard.edu

1. Sebastiani, P. *et al. Science* http://dx.doi.org/10.1126/science.1190532 (2010).
2. Sebastiani, P. *et al. Science* **333,** 404 (2011).
3. Sebastiani, P. *et al. PLoS ONE* **7,** e29848 (2012).
4. Surolia, I. *et al. Nature* **466,** 243–247 (2010).
5. Hunt, K. A. *et al. Nature Genet.* **44,** 3–5 (2012).
6. Li, M. *et al. Science* **333,** 53–58 (2011).
7. Kleinman, C. L. & Majewski, J. *Science* **335,** 1302 (2012).
8. Lin, W. *et al. Science* **335,** 1302 (2012).
9. Pickrell, J. K., Gilad, Y. & Pritchard, J. K. *Science* **335,** 1302 (2012).



LEFT: RICHARD MASSEY; RIGHT: MARTIN KORNMESSER/ESA/HUBBLE

A two-dimensional map (left) of the Universe's mysterious dark matter can be translated into three dimensions (right) for clarity and impact.

VISUALIZATION

Picturing science

Katy Börner weighs up a lavish, lab-friendly guide to transforming dry data into insightful images.

Images of scientific results matter. They support data exploration and communication, and are particularly valuable in the age of big data. They help to transcend disciplinary, cultural and language barriers. Some are truly beautiful, and are displayed in art museums. Others have gone viral on the Internet. A select few — such as Darwin's tree of life or Watson and Crick's DNA structure — have changed our view of the world forever.

Given the importance of imagery in science, it is surprising that few scholars are properly trained in graphic design or data visualization. *Visual Strategies* aims to redress the balance in a format resembling a lavish 'Design 101' textbook — complete with tabs to ease navigation and a plastic cover suited to the wet lab. Felice Frankel and Angela DePace aim to guide scientists and engineers to create graphics that effectively communicate research results. Their case studies gain much from DePace's systems-biology research and Frankel's decades in science photography and image-making, which are showcased in books such as *Envisioning Science: The Design and Craft of the Science Image* (MIT Press, 2002).

Frankel and DePace have identified three main types of visualization: form and structure, process and time, and compare and contrast. The first part of *Visual Strategies* is structured around these categories, and the 18 before-and-after examples within them are among the book's major assets.

Visual Strategies: A Practical Guide to Graphics for Scientists and Engineers
FELICE C. FRANKEL AND ANGELA H. DEPACE
Yale University Press: 2012. 160 pp. £25.00/$35.00

For each example, the answers to five key questions are given — the graphic's intended audience, usage, goal and challenges, as well as suggestions for best approaching its design — along with a list of 'graphical tools' that identifies the approaches used to improve the visualization. The list covers composition, abstraction, colours and layers, as well as refinements used to make the image both insightful, and compelling enough to drive the message home.

A number of case studies and interactive graphics, originally published in leading science journals, are also reproduced, with descriptions and comments by the authors. Interactive graphics that support the book's content can be accessed online, along with the sharing and discussion of graphics (http://visual-strategies.org). The book concludes with a 'visual index' that provides references for all of the examples; the appendix introduces the online forum.

Frankel and DePace's three visualization types make up just one system in a hugely varied field. IBM's ManyEyes website (www-958.ibm.com), for instance, allows users to generate visualizations to reveal relationships between data, compare data values, track rises and falls over time, see parts of a whole, analyse text and generate maps. By contrast, Nathan Yau's *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics* (Wiley, 2011) distinguishes five types of visualization — patterns over time, proportions, relationships, differences and spatial relationships — and also looks at identifying the 'story', handling data and selecting tools.

Beyond identifying and explaining different visualization types, writing a practical guide entails deciding the number and content of examples needed to illustrate workflows and key design decisions, and whether they should be concrete or abstract; specific or general; simple or complex; focused or diverse. In *Visual Strategies*, the before-and-after examples cover many areas of science in great detail — such as the 'quantum corralling' of iron atoms, or the loss of neurons in specific parts of the brain in patients with Alzheimer's disease. One is left with a deep admiration for scientist-designers such

➲ NATURE.COM
For a review of Frankel's work on the nanoscale, see:
go.nature.com/qvaaxl

as Frankel, who within a few weeks or months manage to understand scientific problems and solutions, and 'translate' them visually to communicate important structures and dynamics to a large audience. Although the concrete examples are helpful, their complexity and diversity makes it hard to extract general strategies for effective visualization design. A glossary would have helped to define the terminology used.

As admirable as all these guides — including Frankel and DePace's — are, what is really needed is a general theory for the effective teaching and usage of graphical methods. It might be time for another round of US National Science Foundation workshops or Image and Meaning events (www.imageandmeaning.org), which bring together scientists, writers and visual communicators to develop and share improved methods of communicating scientific results through images and visual representations.

Such a theory would build on a range of sources, such as William Playfair's 1786 publication *The Commercial and Political Atlas*, Jacques Bertin's 1967 *Semiology of Graphics: Diagrams, Networks, Maps*, John Turkey's practical epistemology, Bill Cleveland's combination of statistical and experimental evidence, Edward Tufte's many examples of "beautiful evidence", and Leland Wilkinson's *The Grammar of Graphics* (Springer, 2000). It would draw from psychology, cartography, statistics and other sciences that use data analysis, visualization, graphic design and illustration.

"Given the importance of imagery in science, it is surprising that few scholars are trained in graphic design."

That theory might in turn become the basis for a 'visualization design cookbook'. Taking the reader's information needs as input, this would output a 'shopping list' of relevant data sets, tools and workflows; detail and exemplify each data analysis and visualization step; show pictures of the anticipated end result; point out major challenges; and provide suggestions on how best to meet them. Then, all you would have to do is find the best 'recipe', follow the instructions, make and interpret the visualizations — and the power would be with you. ■

Katy Börner *is the Victor H. Yngve Professor of Information Science at Indiana University Bloomington and curator of the international* Places & Spaces: Mapping Science *exhibition.*
e-mail: katy@indiana.edu

RENAISSANCE ART

Puzzles beneath Dürer's paint

Alison Abbott assesses an infrared analysis that could point to the artist's real ambitions.

Albrecht Dürer (1471–1528) — painter, print-maker and author of books on mathematics, perspective and other aspects of applied science — has been a magnet for art historians since his youth. Nearly 11,000 books and articles have been written about him and his works. Is there anything left to say?

Yes, thanks to a three-year grant from the German government, which has funded an unusual collaboration between art historians and scientists at the German National Museum in the artist's hometown, Nuremberg. The scientists used infrared reflectography systematically to look beneath the surface of Dürer's paintings around the world, and revealed that the levels of detail in the underdrawings vary, both within and between paintings. Their insights now inform the museum's latest blockbuster exhibition, *The Early Dürer*. The art historians involved speculate that the Renaissance man's main ambition was to be an art theorist.

The son of an immigrant goldsmith, Dürer quickly emerged as a force to rival his contemporary, Leonardo da Vinci. Venetian painter Giovanni Bellini is said — perhaps apocryphally — to have asked Dürer for the brush with which he crafted his startlingly realistic beards, with their individual hairs. Dürer gave him an ordinary brush; it was his skill that was extraordinary.

But geographically distanced from the major centres of European art in Italy and the Netherlands, how did Dürer develop his painterly skills? Was he primarily a businessman generating brilliant prints for cash, or a selfless genius in the service of art? Was this prolific self-portraitist, who signed his works with a curious monogram, really an egomaniac? The research project provided no conclusive answers to the many open questions, but did generate a horde of facts for art historians to chew over.

In 2009, the scientists packed up their mobile infrared-reflectography machine for a tour of Dürer's early paintings in 20 museums and private collections in Europe and the United States. Until then, art historians had made much of the scraps of underdrawing revealed by thinning patches of paint or partial infrared analysis. But this systematic analysis of 45 paintings shows that the underdrawings don't actually reveal much about Dürer's artistic development.

The Early Dürer
Germanic National Museum, Nuremberg. Until 2 September 2012.

AKG-IMAGES

Albrecht Dürer's *Self-Portrait* (1500) was a private attempt to stretch the limits of technique.

What they do reveal is the extreme effort that Dürer put into the works he did not intend to sell — such as his self-portraits — in comparison with the works that he did under commercial contract. The detailed underdrawings of his self-portraits indicate his attempts to push the boundaries of artistic technique. The curators speculate that the self-portraits — the last completed when he was just 28 — represent his early, innovative striving for perfection. He may, they posit, have generated his large body of commercial artworks, with their varying levels of workmanship, to gain economic freedom to carry through his great ambitions in art theory. Dürer had planned a series of ten volumes on art theory, but died before he could finish the task. ■

Alison Abbott *is* Nature's *Senior European correspondent.*

Correspondence

Mandate ethics methods in papers

We call for a commitment to ethical reproducibility in biomedical research that would require detailed reporting of research-ethics methods in published scientific reports. We see this as an essential complement to scientific reproducibility.

The current ethics-review system does not make information openly available (E. J. Emanuel *et al. Ann. Intern. Med.* **141,** 282–291; 2004). Furthermore, the only explicit guidelines for ethics reporting are those issued by the International Committee of Medical Journal Editors in its 'Uniform Requirements for Manuscripts'. Few papers actually follow these (S. Schroter *et al. J. Med. Ethics* **32,** 718–723; 2006).

Incorporating details of research-ethics methods into biomedical papers as standard procedure will promote efficient, cost-effective ethics practices and improve the credibility and accountability of research.

James A. Anderson, Marleen Eijkholt, Judy Illes *National Core for Neuroethics, University of British Columbia, Vancouver, Canada.*
jilles@mail.ubc.ca

Patients' research priorities get funded

We are pleased to report evidence of a sea change in how medical research is supported in the United Kingdom. A major clinical research funder is commissioning studies of four of the top ten schizophrenia research questions prioritized by patients, carers and clinicians last year using the James Lind Alliance priority-setting process (see go.nature.com/u2cqcy and *Nature* **474,** 277–278; 2011).

The Health Technology Assessment programme of the National Institute for Health Research funds rigorous and independent research into the effectiveness of health-care technologies relevant to the UK National Health Service (NHS). The schizophrenia questions that the programme will investigate are: how to improve management of the weight gain and sexual dysfunction that is associated with medication; how to manage people whose schizophrenia is unresponsive to treatment; and how to recognize early signs of relapse.

James Lind Alliance partnerships ensure that the views of patients and others outside academia and industry are heard when setting research priorities. From April 2013, the alliance will be coordinated from within the NHS (see go.nature.com/twhvxz). This is an important further step towards incorporating patient, carer and clinician priorities in the wider activities of policy-makers and research funders.

Keith Lloyd, Jo White *Swansea University, Swansea, UK.*
k.r.lloyd@swansea.ac.uk
Iain Chalmers *James Lind Library, The James Lind Initiative, Oxford, UK.*

More medieval clues to cosmic-ray event

Jonathon Allen quotes an entry in the *Anglo-Saxon Chronicle* that might account for the increased cosmic-ray flux in AD 774–775 (*Nature* **486,** 473; 2012). Other medieval texts recall another celestial phenomenon from around the same time, which may or may not be pertinent.

In the context of Charlemagne's campaign against the Saxons, the annals of the monastery of Lorsch, Germany (*Annales Laurissenses*), mention an image witnessed in AD 776 as "two shields burning with red colour and moving above the church itself". The *Chronicon* of Sigebert of Gembloux notes that "when the Saxons besieged the castle of Heresburch, the glory of God appeared to all, surely as two shields burning with the colour of blood and making certain motions through the air, as if at war". The phenomenon seems to have been observed during the day, suggesting that it was very bright if indeed it was a cosmic event.

The *Anglo-Saxon Chronicle* also describes a heavenly red crucifix, a colour that is a traditional motif of battle-related portents. The date disparity between the shield sighting and the AD 774–775 event might be explained by an extended period of auroral activity. Also, the *Anglo-Saxon Chronicle* entry is linked to the Battle of Otford, thought to have occurred in AD 776.

Gary W. Gibbons *University of Cambridge, Cambridge, UK.*
Marcus C. Werner *University of Tokyo, Kashiwa, Japan.*
marcus.werner@ipmu.jp

Political backing to save the Baltic Sea

As Daniel Conley points out, reducing nutrient input is the most important measure for rescuing the Baltic Sea from eutrophication (*Nature* **486,** 463–464; 2012). But its restoration also depends strongly on political support. With ecological models indicating that recovery could take 50–100 years, such support must continue for decades. This will happen only if accompanied by evidence of steady progress.

The Baltic Sea's nutrient load can be reduced by improving agricultural practices, but huge nutrient stores in the soils will slow progress. Yet the European Union wants high water quality to be achieved by 2015 in coastal waters (Water Framework Directive) and by 2020 in open sea areas (Marine Directive).

Patience is not enough: we need ways to accelerate recovery and early demonstrations of improved water quality, at least in some enclosed coastal waters. There has been some progress near Stockholm after upgrades to sewage treatment, even as conditions deteriorated in the open Baltic Sea.

The foundation BalticSea2020 (www.balticsea2020.org/english) will reduce land-based nutrient inputs to a eutrophic bay in the Stockholm archipelago and use aluminium chloride to increase phosphorus binding by the bay's hypoxic sediments. This will reduce nutrient levels and algal production within years. Submerged vegetation and spawning sites for predatory fish will also be restored.

Such timely local recoveries could help to sustain political support for nutrient reductions for long enough to save the Baltic Sea.

Ragnar Elmgren* *Stockholm University, Sweden.*
ragnar.elmgren@ecology.su.se
**On behalf of 4 co-authors (for a full list, see go.nature.com/iefbcy).*

Open access: hard on lone authors

Nowhere in your discussion on the future of author payments for open-access publication (*Nature* **486,** 439; 2012) do you mention the predicament of the independent researcher or, for that matter, the scholar who is not funded by grants. I trust that the authors of the Finch report have borne this in mind.

Otherwise, the paywalls that prevent free access to knowledge for those who are not members of a university or other academic library will merely be replaced by article-publishing charges that prevent them from making a contribution.

Christopher Smith *Birmingham, UK. c.u.m.smith@googlemail.com*

CONTRIBUTIONS
Correspondence may be sent to **correspondence@nature.com** after consulting the author guidelines at **go.nature.com/cmchno**.

A. TREJO/SHUTTERSTOCK

The Brandenburg Gate in Berlin presides over a city that hosts three universities deemed excellent in a federal initiative.

GERMANY

Excellence revisited

The second phase of the German Excellence Initiative is helping scientists and universities. But sustaining its gains in the long term could be a challenge.

BY QUIRIN SCHIERMEIER

Peter Fierlinger talks with disarming modesty about his achievements and aspirations. It is easy to forget that the 34-year-old neutron researcher can already look back on a remarkable career exploring hot topics in fundamental physics. "A wunderkind? Heaven forbid, I was rather a late bloomer," he says. "I guess I was in the right place at the right time."

For Fierlinger, the right place was the Technical University of Munich (TUM) in Germany, where he leads a high-profile research group studying the origin and structure of the Universe. The right time was four years ago, when the TUM physics department offered him a tenure-track position. The department was growing after an injection of money from Germany's Excellence Initiative, a government-funded €4.6-billion (US$5.6-billion) scheme to make Germany's university research system more attractive to international talent and more competitive with systems in the United Kingdom, the United States and elsewhere. The German science ministry announced awards in the initiative's second phase last month, and many more researchers stand to benefit.

At a time when recession and debt crises have prompted other countries to freeze or cut public spending on science, the awards have helped to buoy German research. The programme has created 17,000 science jobs at all levels, says the German Research Foundation (DFG), the country's main funding agency. Supporting postdocs has been a particular goal. And much of the money is being used for hiring and to attract foreign talent. But the initiative is set to end in 2017. To ensure that its impact lasts and that job gains are permanent, the government and universities will have to find ways to keep making big investments.

The TUM was one of the winners in the Excellence Initiative's first round, in 2006. Alongside the Ludwig Maximilian ▶

University (LMU) in Munich, Karlsruhe University (now part of the Karlsruhe Institute of Technology), the Free University of Berlin and the universities of Konstanz, Freiburg, Heidelberg, Göttingen and Aachen, the TUM also received an award for 'future concepts', to fund a campus-wide programme of development and improvement. The initiative also funded graduate schools and large-scale research collaborations — 'excellence clusters' — across all disciplines. A substantial portion of the money went towards salaries and recruiting.

The second phase of awards spreads cash across 99 separate projects at almost 40 institutions (see 'Excellence all round'), including large-scale projects at universities from Bremen in the north to Konstanz in the south — each offering job opportunities. There are now 43 clusters of excellence: 31 carried forward from the first phase, and 12 new ones, with projects including the Immune Sensory System in Bonn, BrainTools in Freiburg and Systems Neurology at the LMU.

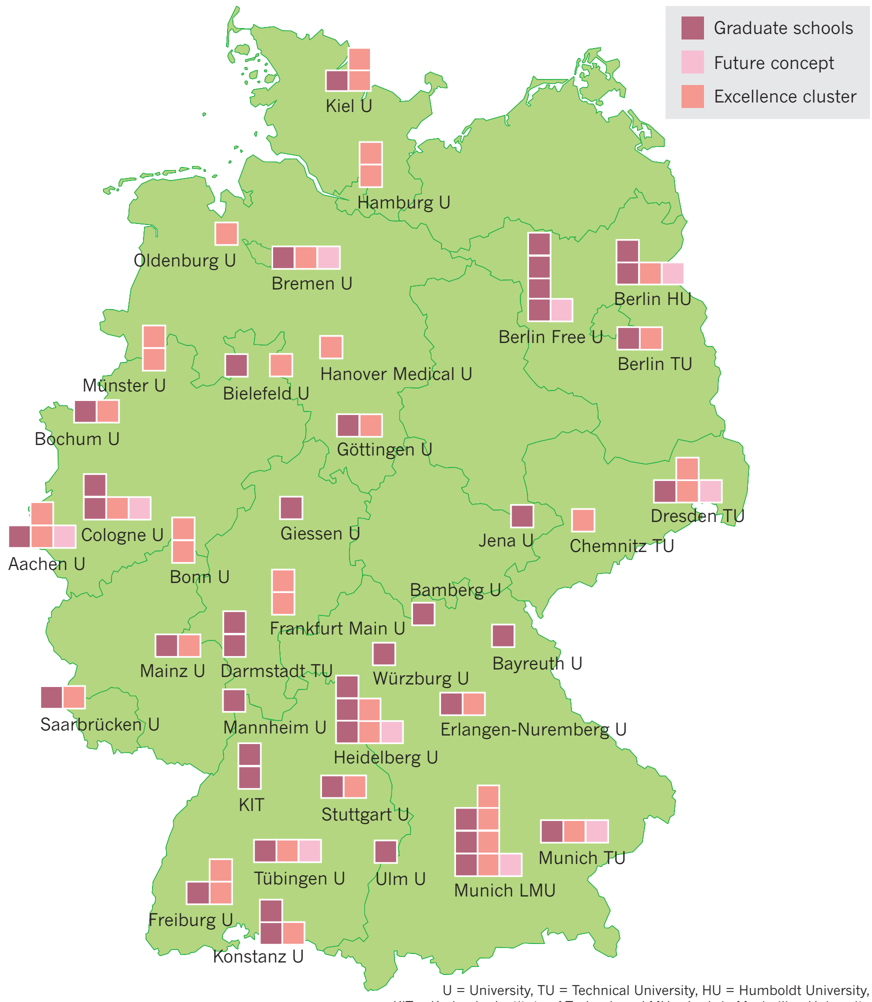
Future-concepts awards at six institutions have been renewed, and new awards have gone to a further five — the Technical University of Dresden, the Humboldt University in Berlin and the universities of Bremen, Cologne and Tübingen. Each institution gets an extra €12.5 million per year; only universities that have at least one initiative-funded graduate school and one excellence cluster are eligible.

The financial bonus from the future-concepts award may not in itself be exceedingly generous on the scale of funds available in some other countries. But the competition as a whole has rejuvenated German science by, among other things, stimulating new research proposals and encouraging scientists to increase collaboration across departments and disciplines.

EXCELLENCE ALL ROUND

This year, the second phase of Germany's Excellence Initiative granted federal funding to research projects ('excellence clusters', not all shown), graduate programmes and campus-wide improvement strategies ('future concepts') at universities around the country.

SOURCE: DFG



U = University, TU = Technical University, HU = Humboldt University, KIT = Karlsruhe Institute of Technology, LMU = Ludwig Maximilian University

WINNING FORMULA?

When it became law in 2005, the initiative had its share of critics. Some argued that the German system was better left alone, or that elite institutions should be allowed to evolve on their own without orchestrated government support, says Jan-Hendrik Olbertz, president of the Humboldt University. The scheme's excessive paperwork was an irritant for many; scientists at candidate universities had to contribute to applications. But Olbertz says that this "exercise in institutional governance" has forced Germany's entire academic elite to think hard about which parts of its science and higher-education system work well, and which do not.

German universities have not historically tended to have US- and UK-style postdoc positions, and it has become apparent that this is a severe impediment to recruiting international talent. "We're of similar size but we have five times fewer postdocs than the University of Cambridge in Britain," says Wolfgang Herrmann, president of the TUM. "If we want to offer young scientists better career prospects, and we certainly do, this must urgently change."

Furthermore, in the German system, scientists have traditionally been promoted on the basis of habilitation: a qualification after the doctorate, which requires a second dissertation. Habilitation is no longer mandatory (see *Nature* **450**, 452–453; 2007), but still exists, and under the rule of *Hausberufungsverbot*, academics cannot become assistant professors at the university at which they were habilitated. There is also much less emphasis on a cumulative publication record in Germany than in the United States, and less of a defined career structure.

Under the Excellence Initiative, some universities are shaking up that system. The TUM, for example, has ceased to appoint junior professors, who usually cannot earn tenure at their host university. Instead, it has set up a tenure-track programme that provides a seamless path from fixed-term postdoc or assistant-professor positions to tenured associate professorships and full professorships.

The initiative has also helped to spur more careful thinking on early-career progression and mentoring programmes. In one scheme at the TUM, young scientists will be coached in personnel management.

The DFG says that the initiative has attracted talent to Germany: between 2005 and 2009, the number of foreign scientists at German universities and research institutes increased by one-third, and many German-born scientists working abroad have returned. Several universities have stepped up recruitment — the TUM, for example, opened a liaison office in São Paulo, Brazil, this year.

SUSTAINABLE SUPPORT

The TUM's streamlined recruitment policies, which now include considerably more room for negotiating salaries, enabled Fierlinger's return to Europe from Stanford University in California, where he had spent two years as a postdoc. "I wouldn't have thought it in the first place, but the Excellence Initiative has really become a self-fulfilling prophecy," says Fierlinger, who received tenure in 2010. "Germany has become a much more attractive science destination."

But the initiative could also mean that

Germany will soon face a surplus of postdocs or young scientists — already a problem in countries such as the United States. The number of postdocs alone suggests that many of the young scientists who have been recruited in recent years will not find scientific jobs in Germany, especially once the initiative's funding has run out.

There are efforts under way to support non-academic career paths. Many of the universities that have received initiative money have set up scientific-career training programmes. Young scientists can also get extra guidance in subjects ranging from patent management to business administration, notes Oliver Baron, managing director of the Center for Integrated Protein Science Munich (CIPSM) at the LMU — one of the research clusters selected for extra funding in the initiative's second phase.

Still, it is unclear what will happen after the initiative ends. Universities might suffer, particularly in poorer states such as Berlin and Brandenburg. At risk is the initiative's main goal — creating universities of international repute. And if there is no follow-up funding, scientists at centres such as the CIPSM or the Courant centres, set up at the University of Göttingen to promote independent research, could lose their jobs.

At universities such as Göttingen, administrators are starting to look for options to preserve what the initiative has helped them to build up — whether that is funding for hiring, collaborative projects or graduate programmes. And federal minister of education and research Annette Schavan has proposed that, in the interest of fairness, universities that won funding only in the second phase of the scheme should be granted another round of money.

Schavan also hopes that a planned amendment to Germany's constitutional law will help. The law, soon to be considered by parliament, would allow the federal government to co-finance universities permanently, rather than leaving it to state governments (see *Nature* **483,** 245–246; 2012). The change could pave the way for a federally funded university modelled on the Swiss federal institutes of technology in Zurich and Lausanne.

While administrators and policy-makers ponder the future of funding, Fierlinger is more concerned with truly universal questions. Five years of hard work, he hopes, might turn a fundamental property of neutrons into a gauge of how matter and antimatter are distributed in the cosmos. Administrators and scientists hope that another five years of the Excellence Initiative will be enough to establish lasting change and benefits. ■

Quirin Schiermeier *is* Nature*'s Germany correspondent.*

TURNING POINT
Robert Gregg

Robert Gregg works at the boundary between engineering and medicine. A postdoc at the Center for Bionic Medicine at the Rehabilitation Institute of Chicago in Illinois, Gregg received a 2012 Burroughs Wellcome Fund Career Award at the Scientific Interface in May for his work developing robotic control systems for prostheses. He explains how he will use the funds when he joins the bioengineering and mechanical engineering departments at the University of Texas, Dallas, next year.

A. E. SCHULTZ

Which came first — an interest in robots or in prosthetics?

I did a degree in electrical engineering at the University of California, Berkeley, and had an internship studying control principles of robotic walking. That sparked my interest in robots. I loved the Terminator movies and, at first, I was following the 'cool factor'. I went to graduate school at the University of Illinois at Urbana-Champaign because it had a leading programme on control theory about the intelligence behind machines. I never intended to build killer robots, but I didn't understand until later how this research could help people.

How did you realize the potential for medical applications?

In my last year at graduate school, I was thinking about where I wanted to take my career. Then my father was diagnosed with heart failure. I had to take 2–3 months off while he waited for a transplant. The doctors told me about a surgical robot they were using; that helped me to realize that people could benefit from my research. So I decided to pursue biomedical engineering.

How did you get your postdoc?

I won one of five Engineering into Medicine fellowships at the Northwestern University Clinical and Translational Science Institute (NUCATS) in Chicago. It was a one-time opportunity funded by the American Recovery and Reinvestment Act of 2009, and happened at the right place and time for me. NUCATS was looking for people who could translate engineering principles into medical or biological research. It was the biggest turning point of my career: I was able to apply ideas from robotics and control theory to improve the performance of prosthetic devices. I'm experimenting with a prosthetic-leg controller that moves joints based on measurements of pressure to the sole of the foot. The wars in Iraq and Afghanistan have brought attention to the needs of amputees, but amputations are necessitated most frequently by disease, especially cardiovascular disease and diabetes.

What was most difficult about switching research cultures?

Moving into biomedicine was a leap, because no one in the field knew of my engineering research. I had no reputation in prosthetics and no knowledge of it, really. But the most difficult part was learning how to communicate with clinicians. The same words can mean completely different things. For example, 'control' means the intelligence behind a machine to an engineer, whereas clinicians use it to describe study patients who do not receive the experimental treatment.

Was the search for a job after your postdoc difficult?

This spring was a whirlwind. The market is rough — I applied for 35 faculty positions and it looked like nothing would pan out. Then I was approached by the University of Texas, Dallas. I also learned that I was a finalist for the Burroughs Wellcome award, which comes with US$500,000 to finish a postdoc and start a lab. In a two-day period, I got the award and the offer to start at Dallas in June 2013.

How does the award help your career?

It will let me hire personnel in biomedical and mechanical engineering, and to start off with expensive research. I am eyeing a treadmill that costs more than some houses. It measures the forces that human feet transfer to the ground, to analyse the performances of prostheses. I think of my lab as a start-up company — I have investors; now I have to prove that my vision for robot-assisted walking can be achieved. ■

INTERVIEW BY VIRGINIA GEWIN

Atmospheric oxygenation and volcanism

ARISING FROM F. Gaillard, B. Scaillet & N. T. Arndt *Nature* **478,** 229–232 (2011)

Around 2.5 billion years ago, Earth's atmosphere turned from anoxic to oxic in what is known as the Great Oxidation Event. Gaillard *et al.*[1] suggest that this oxygenation was caused by the emergence of the continents and a shift in volcanism from predominantly submarine to primarily subaerial conditions. Because the ratio of volcanic SO_2 to H_2S in their model increases with this shift, they argue that the atmosphere became more oxidized. But their model also predicts that outgassing of CO_2 decreases at the lower pressure of continental volcanism, and that this should act against atmospheric oxygenation because CO_2 is the substrate for the production of oxygen from photosynthesis. Hence, their mechanism may not trigger a rise in atmospheric O_2.

There are three problems to consider. First, Gaillard *et al.*[1] argue that the release of SO_2 comes at the expense of H_2S, according to:

$$H_2S + 2H_2O \rightarrow SO_2 + 3H_2 \qquad (1)$$

But the extra H_2 released by this reaction exactly compensates for the switch from H_2S to SO_2. The mixture of SO_2 and H_2 has the same redox potential as H_2S and water[2]. Fortunately, what happens in their model is different. The H_2S to SO_2 ratio is buffered by magma, which has a fixed oxidation state, described in terms of a fixed oxygen fugacity. The relevant reaction is:

$$H_2S + \text{`}O_2\text{'} \rightarrow SO_2 + H_2 \qquad (2)$$

where 'O_2' represents oxidants (mostly ferric iron) in the melt. Reaction (2) produces SO_2 and H_2 in equimolar proportions, in agreement with results shown in figure 2 of ref. 1. When SO_2 is reduced to pyrite, FeS_2, it consumes hydrogen as follows:

$$4\,SO_2 + 10\,H_2 + 2\,FeO \rightarrow 2\,FeS_2 + 10\,H_2O \qquad (3)$$

Because reaction (3) consumes more hydrogen per mole of SO_2 than is produced by reaction (2), the net effect is to oxidize the atmosphere. Thus, this part of their proposed atmospheric oxidation mechanism appears to work, but not for the reasons they stated.

The second issue is that sulphur-containing gas fluxes are not the only ones that control O_2. CO_2 is crucial because some CO_2 is reduced to organic carbon, CH_2O, which is then buried. A mole of photosynthetic O_2 is released for every mole of buried CH_2O. However, O_2 can be lost in atmospheric reaction with H_2, so the net effect is:

$$CO_2 + 2H_2 \rightarrow CH_2O + H_2O \qquad (4)$$

The carbon isotope record suggests that $\sim$20% of outgassed CO_2 is reduced and buried in this manner[3,4]. Holland[4] quantified the effect of volcanic gases on the hydrogen budget by defining a parameter denoted *f*. Gases for which $f > 1$ can reduce 20% of the CO_2 to organic matter and all of the SO_2 to pyrite, leaving excess hydrogen to support an anoxic atmosphere; gases with $f < 1$ contain too little hydrogen to do this. Atmospheric O_2 'wins' over hydrogen when *f* drops below unity. This analysis neglects other sources of hydrogen that might have been important (for example, serpentinization of ultramafic rocks on the continents and seafloor, or precipitation of ferric iron in banded iron formations), but it is useful for analysing the volcanic gases of ref. 1.

In the model of Gaillard *et al.*[1], the proportion of CO_2 in released volcanic gases decreases by a factor of $\sim$3 when volcanoes switch from submarine to subaerial conditions because the pressure decreases from 100 bar to 1 bar. This change should tend to oppose atmospheric oxygenation, as less organic matter is buried (given a constant burial fraction), and thus less H_2 is consumed. To estimate the net effect on atmospheric redox balance, we calculated *f* values for cases listed in the Supplementary Information of ref. 1. Gaillard *et al.*[1] include S_2 as a species, which requires us to modify Holland's *f* parameter, as follows:

$$f = \frac{m(H_2) + 0.6m(CO) - 0.4m(CO_2) + 3m(H_2S) + 4m(S_2)}{3.5[m(SO_2) + m(H_2S) + 2m(S_2)]} + \frac{1}{3.5}$$

Here, $m(i)$ is the mole fraction of species i in the released gases. As shown in Table 1, f is less than unity in the cases shown in Supplementary Tables 1 and 2 of Gaillard *et al.*[1], and so would not allow a reduced atmosphere before 2.4 Gyr ago. This means that no rise of O_2 is possible because the initial atmosphere is already oxidized. This could be because other hydrogen sources are neglected, as mentioned above, or because model parameters—for example, the outgassing temperature—are incorrect. In any case, the model, as presented in ref. 1, appears incapable of triggering a rise in atmospheric oxygen if Holland's method of analysis is correct.

Last, Gaillard *et al.*[1] assume that gases released from a submarine magma at 1,300 °C do not re-equilibrate with the surrounding basalts at lower temperatures as they emerge from depth; this assumption is at odds with previous work[5,6]. Such re-equilibration might have allowed Archaean volcanic gases to be more reduced; hence, a switch to subaerial volcanism around 2.5 Gyr ago might then result in atmospheric oxidation. But this would be a different oxidation mechanism from the one described in ref. 1. There is also a broader caveat. The reduced oxidation state of volcanic gases comes at the expense of oxidizing the source rocks; thus, erosion and subsequent reduction of the source rocks may cancel out surficial redox change in the long term.

Table 1 | *f* values for different cases from ref. 1

Case	Pressure (bar)	
	1	100
Supplementary Table 1	0.60	0.62
Supplementary Table 2	0.73	0.88

James F. Kasting[1], David C. Catling[2] & Kevin Zahnle[3]
[1]Department of Geosciences, The Pennsylvania State University, 443 Deike, University Park, Pennsylvania 16802, USA.
email: kasting@essc.psu.edu
[2]Department of Earth and Space Sciences, and Astrobiology Program, Box 351310, University of Washington, Seattle, Washington 98195, USA.
[3]NASA Ames Research Center, MS 245-3, Moffett Field, California 94035, USA.

Received 15 November 2011; accepted 17 May 2012.

1. Gaillard, F., Scaillet, B. & Arndt, N. T. Atmospheric oxygenation caused by a change in volcanic degassing pressure. *Nature* **478,** 229–232 (2011).
2. Kasting, J. F. & Brown, L. L. in *The Molecular Origins of Life: Assembling the Pieces of the Puzzle* (ed. Brack, A.) 35–56 (Cambridge Univ. Press, 1998).
3. Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time, *Phil. Trans. R. Soc. Lond. B* **361,** 931–950 (2006).
4. Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66,** 3811–3826 (2002).
5. Kump, L. R. & Seyfried, W. E. Hydrothermal Fe fluxes during the Precambrian: effect of low oceanic sulfate concentrations and low hydrostatic pressure on the composition of black smokers. *Earth Planet. Sci. Lett.* **235,** 654–662 (2005).
6. Kump, L. R. & Barley, M. E. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* **448,** 1033–1036 (2007).

Author Contributions All three authors contributed equally to this Comment.

Competing Financial Interests Declared none.

doi:10.1038/nature11274

Gaillard *et al.* reply

REPLYING TO J. F. Kasting, D. C. Catling & K. Zahnle *Nature* **487,** http://dx.doi.org/10.1038/nature11274 (2012)

Kasting *et al.*[1] question the model of ref. 2, in which we suggest that the oxygenation of the atmosphere, around 2.45 Gyr ago, was promoted by the emergence of subaerial volcanism, producing volcanic gases with much more elevated SO_2/H_2S ratios than submarine volcanism.

Kasting *et al.*[1] claim that the enhanced SO_2/H_2S ratio in subaerial volcanic gases was accompanied by enhanced H_2 production, which may limit the oxidative capacity of emitted gases. This is only partly correct, because enhanced SO_2 also derives from the reaction[2,3]:

$$S^{2-}(\text{melt}) + 3Fe_2O_3\,(\text{melt}) \rightarrow SO_2 + 6FeO\,(\text{melt}) + O^{2-}\,(\text{melt})$$

which implies that subaerial degassing extracts more oxygen from the melt than submarine degassing. The oxygen reservoir of the melt, a fundamental aspect of our model[3] that has so far not been taken into account, implies that more oxygen was therefore degassed as subaerial volcanism became abundant at about 2.7 Gyr ago.

Also, Kasting *et al.*[1] argue that the amount of outgassed CO_2 decreases by a factor of 3 as venting pressure decreases from 100 bar to 1 bar, which should limit production of organic carbon (CH_2O) and thereby limit the associated consumption[4] of atmospheric H_2. However, although the molar fraction of CO_2 in the gas decreases, the flux of CO_2 into the atmosphere is unchanged between 100 and 1 bar venting pressures[2], owing to the exceedingly low solubility of CO_2 in silicate melt in this pressure range (unlike the case for sulphur).

The *f* parameter of Holland[4] is used by Kasting *et al.*[1] to evaluate how much H_2 is consumed to reduce volcanic CO_2 into organic matter and SO_2 to pyrite. According to Kasting *et al.*[1], as pressure decreases, the *f* values of our calculated gas compositions indeed decrease (that is, their reducing power decreases, as required), but do not reach low enough values to drive the atmosphere to oxidizing conditions. However, the calculation of *f* is based on the way H_2S is produced or consumed in volcanic gases: Holland[4] first considered decomposition of H_2S during cooling, which is equivalent to production of H_2 (hence the $+3m(H_2S)$ term in the *f* equation). The more recent analysis[5] by Holland considers instead that H_2S is the product of reaction between SO_2 and H_2 during cooling, a H_2-consuming reaction ($-3m(H_2S)$ in the *f* equation).

Conventionally, about 20% of volcanic CO_2 is consumed to produce organic matter[4]. Any variation of the amount severely affects the results of calculations made using the *f* equation, highlighting the difficulties in using it as to determine the oxidative capacity of volcanic gases. Holland's more recent analysis[5] of the causes of oxygenation suggests that oxidation was due to an increase in CO_2 and SO_2 volcanic fluxes, which is what our model predicts as volcanism changed from quasi-exclusively-submarine to partially subaerial. At this point, we stress that our model[2] not only describes an increase in the oxidative capacity of volcanic gas but also a chain reaction likely to facilitate atmospheric oxygenation. Of prime importance are the sulphate reduction processes, which should have been exacerbated by elevated volcanic SO_2 emissions. Biological sulphate reduction transforms sedimentary organic carbon into CO_2, which results in oxygen production[6]. In parallel, hydrothermal sulphate reduction, which decreases the reducing potential of hydrothermal fluids and fixes hydrothermal ferrous iron as pyrite, also contributed to atmospheric oxygenation[4]. All these reaction paths are not included in Holland's *f* factor, whereas they were certainly involved in the Great Oxidation Event.

We agree with the final recommendation of Kasting *et al.*[1] that both volcanic gases and hydrothermal fluids should be considered in models of the Great Oxidation Event. However, whereas we accept that thick Archaean oceanic crust was on average more mafic than younger crust, the uppermost layers—those most susceptible to hydrothermal alteration—would have consisted of olivine-poor basalt. In both modern oceanic plateaus and presumably in Archaean oceanic crust, parental picritic magma differentiates, leaving olivine cumulates at the Moho and erupting relatively evolved lava[7]. Basalt with little to no olivine is the dominant component of the upper parts of both modern oceanic plateaus and Archaean greenstone belts[7,8]. These rocks are not susceptible to serpentinization; therefore little H_2 would have been produced during their hydrothermal alteration, and its impact on the atmospheric oxygenation should not have been as important as claimed by Kasting *et al.*[1].

Last, Kasting *et al.*[1] expressed concern about low-temperature re-equilibration processes between volcanic gases and basalts that were not considered by us[2]. We answer that this comment seems to re-introduce confusion between volcanic gas inputs (from mantle to exosphere) and hydrothermal recycling (seawater that reacts with basalts) that may arise from a misinterpretation of ref. 9. Volcanic degassing and hydrothermal emissions are two fundamentally distinct processes, which not only differ in temperature, but chiefly differ in their source (igneous input versus surficial recyling).

Fabrice Gaillard[1], Bruno Scaillet[1] & Nicholas T. Arndt[2]
[1]Institut des Sciences de la Terre d'Orléans, CNRS-INSU/Université d'Orléans/Université de Tours, 1a rue de la Férollerie 45071, Orléans cedex 2, France.
[2]ISTerre, Université Joseph Fourier de Grenoble, CNRS, 1381 rue de la Piscine, 38400 Saint Martin d'Hères, France.
email: fabrice.gaillard@cnrs-orleans.fr

1. Kasting, J. F. Catling, D. C. & Zahnle, K. Atmospheric oxygenation and volcanism. *Nature* **487,** http://dx.doi.org/10.1038/nature11274 (2012).
2. Gaillard, F., Scaillet, B. & Arndt, N. T. Atmospheric oxygenation caused by a change in volcanic degassing pressure. *Nature* **478,** 229–232 (2011).
3. Gaillard, F. & Scaillet, B. The sulfur content of volcanic gases on Mars. *Earth Planet. Sci. Lett.* **279,** 34–43 (2009).
4. Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66,** 3811–3826 (2002).
5. Holland, H. D. Why the atmosphere became oxygenated: a proposal. *Geochim. Cosmochim. Acta* **73,** 5241–5255 (2009).
6. Lyons, T. W. & Gill, B. C. Ancient sulfur cycling and oxygenation of the early biosphere. *Elements* **6,** 93–99 (2010).
7. Fitton, J. G., Mahoney, J. J., Wallace, P. J. & Saunders, A. D. (eds) *Origin and Evolution of the Ontong Java Plateau* (Geological Society of London, 2004).
8. de Wit, M. J. & Ashwal, L. D. *Greenstone Belts* (Oxford Scientific, 1997).
9. Kump, L. R. & Barley, M. E. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* **448,** 1033–1036 (2007).

doi:10.1038/nature11275

Atmospheric oxygenation and volcanism

ARISING FROM F. Gaillard, B. Scaillet & N. T. Arndt *Nature* **478,** 229–232 (2011)

Around 2.5 billion years ago, Earth's atmosphere turned from anoxic to oxic in what is known as the Great Oxidation Event. Gaillard *et al.*[1] suggest that this oxygenation was caused by the emergence of the continents and a shift in volcanism from predominantly submarine to primarily subaerial conditions. Because the ratio of volcanic SO_2 to H_2S in their model increases with this shift, they argue that the atmosphere became more oxidized. But their model also predicts that outgassing of CO_2 decreases at the lower pressure of continental volcanism, and that this should act against atmospheric oxygenation because CO_2 is the substrate for the production of oxygen from photosynthesis. Hence, their mechanism may not trigger a rise in atmospheric O_2.

There are three problems to consider. First, Gaillard *et al.*[1] argue that the release of SO_2 comes at the expense of H_2S, according to:

$$H_2S + 2H_2O \rightarrow SO_2 + 3H_2 \quad (1)$$

But the extra H_2 released by this reaction exactly compensates for the switch from H_2S to SO_2. The mixture of SO_2 and H_2 has the same redox potential as H_2S and water[2]. Fortunately, what happens in their model is different. The H_2S to SO_2 ratio is buffered by magma, which has a fixed oxidation state, described in terms of a fixed oxygen fugacity. The relevant reaction is:

$$H_2S + \text{‘}O_2\text{’} \rightarrow SO_2 + H_2 \quad (2)$$

where ‘O_2’ represents oxidants (mostly ferric iron) in the melt. Reaction (2) produces SO_2 and H_2 in equimolar proportions, in agreement with results shown in figure 2 of ref. 1. When SO_2 is reduced to pyrite, FeS_2, it consumes hydrogen as follows:

$$4\, SO_2 + 10\, H_2 + 2\, FeO \rightarrow 2\, FeS_2 + 10\, H_2O \quad (3)$$

Because reaction (3) consumes more hydrogen per mole of SO_2 than is produced by reaction (2), the net effect is to oxidize the atmosphere. Thus, this part of their proposed atmospheric oxidation mechanism appears to work, but not for the reasons they stated.

The second issue is that sulphur-containing gas fluxes are not the only ones that control O_2. CO_2 is crucial because some CO_2 is reduced to organic carbon, CH_2O, which is then buried. A mole of photosynthetic O_2 is released for every mole of buried CH_2O. However, O_2 can be lost in atmospheric reaction with H_2, so the net effect is:

$$CO_2 + 2H_2 \rightarrow CH_2O + H_2O \quad (4)$$

The carbon isotope record suggests that ~20% of outgassed CO_2 is reduced and buried in this manner[3,4]. Holland[4] quantified the effect of volcanic gases on the hydrogen budget by defining a parameter denoted *f*. Gases for which $f > 1$ can reduce 20% of the CO_2 to organic matter and all of the SO_2 to pyrite, leaving excess hydrogen to support an anoxic atmosphere; gases with $f < 1$ contain too little hydrogen to do this. Atmospheric O_2 ‘wins’ over hydrogen when *f* drops below unity. This analysis neglects other sources of hydrogen that might have been important (for example, serpentinization of ultramafic rocks on the continents and seafloor, or precipitation of ferric iron in banded iron formations), but it is useful for analysing the volcanic gases of ref. 1.

In the model of Gaillard *et al.*[1], the proportion of CO_2 in released volcanic gases decreases by a factor of ~3 when volcanoes switch from submarine to subaerial conditions because the pressure decreases from 100 bar to 1 bar. This change should tend to oppose atmospheric oxygenation, as less organic matter is buried (given a constant burial fraction), and thus less H_2 is consumed. To estimate the net effect on atmospheric redox balance, we calculated *f* values for cases listed in the Supplementary Information of ref. 1. Gaillard *et al.*[1] include S_2 as a species, which requires us to modify Holland's *f* parameter, as follows:

$$f = \frac{m(H_2) + 0.6m(CO) - 0.4m(CO_2) + 3m(H_2S) + 4m(S_2)}{3.5[m(SO_2) + m(H_2S) + 2m(S_2)]} + \frac{1}{3.5}$$

Here, $m(i)$ is the mole fraction of species i in the released gases. As shown in Table 1, f is less than unity in the cases shown in Supplementary Tables 1 and 2 of Gaillard *et al.*[1], and so would not allow a reduced atmosphere before 2.4 Gyr ago. This means that no rise of O_2 is possible because the initial atmosphere is already oxidized. This could be because other hydrogen sources are neglected, as mentioned above, or because model parameters—for example, the outgassing temperature—are incorrect. In any case, the model, as presented in ref. 1, appears incapable of triggering a rise in atmospheric oxygen if Holland's method of analysis is correct.

Last, Gaillard *et al.*[1] assume that gases released from a submarine magma at 1,300 °C do not re-equilibrate with the surrounding basalts at lower temperatures as they emerge from depth; this assumption is at odds with previous work[5,6]. Such re-equilibration might have allowed Archaean volcanic gases to be more reduced; hence, a switch to subaerial volcanism around 2.5 Gyr ago might then result in atmospheric oxidation. But this would be a different oxidation mechanism from the one described in ref. 1. There is also a broader caveat. The reduced oxidation state of volcanic gases comes at the expense of oxidizing the source rocks; thus, erosion and subsequent reduction of the source rocks may cancel out surficial redox change in the long term.

Table 1 | *f* values for different cases from ref. 1

Case	Pressure (bar)	
	1	100
Supplementary Table 1	0.60	0.62
Supplementary Table 2	0.73	0.88

James F. Kasting[1], David C. Catling[2] & Kevin Zahnle[3]
[1]Department of Geosciences, The Pennsylvania State University, 443 Deike, University Park, Pennsylvania 16802, USA.
email: kasting@essc.psu.edu
[2]Department of Earth and Space Sciences, and Astrobiology Program, Box 351310, University of Washington, Seattle, Washington 98195, USA.
[3]NASA Ames Research Center, MS 245-3, Moffett Field, California 94035, USA.

Received 15 November 2011; accepted 17 May 2012.

1. Gaillard, F., Scaillet, B. & Arndt, N. T. Atmospheric oxygenation caused by a change in volcanic degassing pressure. *Nature* **478,** 229–232 (2011).
2. Kasting, J. F. & Brown, L. L. in *The Molecular Origins of Life: Assembling the Pieces of the Puzzle* (ed. Brack, A.) 35–56 (Cambridge Univ. Press, 1998).
3. Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time, *Phil. Trans. R. Soc. Lond. B* **361,** 931–950 (2006).
4. Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66,** 3811–3826 (2002).
5. Kump, L. R. & Seyfried, W. E. Hydrothermal Fe fluxes during the Precambrian: effect of low oceanic sulfate concentrations and low hydrostatic pressure on the composition of black smokers. *Earth Planet. Sci. Lett.* **235,** 654–662 (2005).
6. Kump, L. R. & Barley, M. E. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* **448,** 1033–1036 (2007).

Author Contributions All three authors contributed equally to this Comment.

Competing Financial Interests Declared none.

doi:10.1038/nature11274

Gaillard *et al.* reply

REPLYING TO J. F. Kasting, D. C. Catling & K. Zahnle *Nature* **487,** http://dx.doi.org/10.1038/nature11274 (2012)

Kasting *et al.*[1] question the model of ref. 2, in which we suggest that the oxygenation of the atmosphere, around 2.45 Gyr ago, was promoted by the emergence of subaerial volcanism, producing volcanic gases with much more elevated SO_2/H_2S ratios than submarine volcanism.

Kasting *et al.*[1] claim that the enhanced SO_2/H_2S ratio in subaerial volcanic gases was accompanied by enhanced H_2 production, which may limit the oxidative capacity of emitted gases. This is only partly correct, because enhanced SO_2 also derives from the reaction[2,3]:

$$S^{2-}\,(\mathrm{melt}) + 3Fe_2O_3\,(\mathrm{melt}) \rightarrow SO_2 + 6FeO\,(\mathrm{melt}) + O^{2-}\,(\mathrm{melt})$$

which implies that subaerial degassing extracts more oxygen from the melt than submarine degassing. The oxygen reservoir of the melt, a fundamental aspect of our model[3] that has so far not been taken into account, implies that more oxygen was therefore degassed as subaerial volcanism became abundant at about 2.7 Gyr ago.

Also, Kasting *et al.*[1] argue that the amount of outgassed CO_2 decreases by a factor of 3 as venting pressure decreases from 100 bar to 1 bar, which should limit production of organic carbon (CH_2O) and thereby limit the associated consumption[4] of atmospheric H_2. However, although the molar fraction of CO_2 in the gas decreases, the flux of CO_2 into the atmosphere is unchanged between 100 and 1 bar venting pressures[2], owing to the exceedingly low solubility of CO_2 in silicate melt in this pressure range (unlike the case for sulphur).

The *f* parameter of Holland[4] is used by Kasting *et al.*[1] to evaluate how much H_2 is consumed to reduce volcanic CO_2 into organic matter and SO_2 to pyrite. According to Kasting *et al.*[1], as pressure decreases, the *f* values of our calculated gas compositions indeed decrease (that is, their reducing power decreases, as required), but do not reach low enough values to drive the atmosphere to oxidizing conditions. However, the calculation of *f* is based on the way H_2S is produced or consumed in volcanic gases: Holland[4] first considered decomposition of H_2S during cooling, which is equivalent to production of H_2 (hence the $+3m(H_2S)$ term in the *f* equation). The more recent analysis[5] by Holland considers instead that H_2S is the product of reaction between SO_2 and H_2 during cooling, a H_2-consuming reaction ($-3m(H_2S)$ in the *f* equation).

Conventionally, about 20% of volcanic CO_2 is consumed to produce organic matter[4]. Any variation of the amount severely affects the results of calculations made using the *f* equation, highlighting the difficulties in using it as to determine the oxidative capacity of volcanic gases. Holland's more recent analysis[5] of the causes of oxygenation suggests that oxidation was due to an increase in CO_2 and SO_2 volcanic fluxes, which is what our model predicts as volcanism changed from quasi-exclusively-submarine to partially subaerial. At this point, we stress that our model[2] not only describes an increase in the oxidative capacity of volcanic gas but also a chain reaction likely to facilitate atmospheric oxygenation. Of prime importance are the sulphate reduction processes, which should have been exacerbated by elevated volcanic SO_2 emissions. Biological sulphate reduction transforms sedimentary organic carbon into CO_2, which results in oxygen production[6]. In parallel, hydrothermal sulphate reduction, which decreases the reducing potential of hydrothermal fluids and fixes hydrothermal ferrous iron as pyrite, also contributed to atmospheric oxygenation[4]. All these reaction paths are not included in Holland's *f* factor, whereas they were certainly involved in the Great Oxidation Event.

We agree with the final recommendation of Kasting *et al.*[1] that both volcanic gases and hydrothermal fluids should be considered in models of the Great Oxidation Event. However, whereas we accept that thick Archaean oceanic crust was on average more mafic than younger crust, the uppermost layers—those most susceptible to hydrothermal alteration—would have consisted of olivine-poor basalt. In both modern oceanic plateaus and presumably in Archaean oceanic crust, parental picritic magma differentiates, leaving olivine cumulates at the Moho and erupting relatively evolved lava[7]. Basalt with little to no olivine is the dominant component of the upper parts of both modern oceanic plateaus and Archaean greenstone belts[7,8]. These rocks are not susceptible to serpentinization; therefore little H_2 would have been produced during their hydrothermal alteration, and its impact on the atmospheric oxygenation should not have been as important as claimed by Kasting *et al.*[1].

Last, Kasting *et al.*[1] expressed concern about low-temperature re-equilibration processes between volcanic gases and basalts that were not considered by us[2]. We answer that this comment seems to re-introduce confusion between volcanic gas inputs (from mantle to exosphere) and hydrothermal recycling (seawater that reacts with basalts) that may arise from a misinterpretation of ref. 9. Volcanic degassing and hydrothermal emissions are two fundamentally distinct processes, which not only differ in temperature, but chiefly differ in their source (igneous input versus surficial recyling).

Fabrice Gaillard[1], Bruno Scaillet[1] & Nicholas T. Arndt[2]
[1]Institut des Sciences de la Terre d'Orléans, CNRS-INSU/Université d'Orléans/Université de Tours, 1a rue de la Férollerie 45071, Orléans cedex 2, France.
[2]ISTerre, Université Joseph Fourier de Grenoble, CNRS, 1381 rue de la Piscine, 38400 Saint Martin d'Hères, France.
email: fabrice.gaillard@cnrs-orleans.fr

1. Kasting, J. F. Catling, D. C. & Zahnle, K. Atmospheric oxygenation and volcanism. *Nature* **487,** http://dx.doi.org/10.1038/nature11274 (2012).
2. Gaillard, F., Scaillet, B. & Arndt, N. T. Atmospheric oxygenation caused by a change in volcanic degassing pressure. *Nature* **478,** 229–232 (2011).
3. Gaillard, F. & Scaillet, B. The sulfur content of volcanic gases on Mars. *Earth Planet. Sci. Lett.* **279,** 34–43 (2009).
4. Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66,** 3811–3826 (2002).
5. Holland, H. D. Why the atmosphere became oxygenated: a proposal. *Geochim. Cosmochim. Acta* **73,** 5241–5255 (2009).
6. Lyons, T. W. & Gill, B. C. Ancient sulfur cycling and oxygenation of the early biosphere. *Elements* **6,** 93–99 (2010).
7. Fitton, J. G., Mahoney, J. J., Wallace, P. J. & Saunders, A. D. (eds) *Origin and Evolution of the Ontong Java Plateau* (Geological Society of London, 2004).
8. de Wit, M. J. & Ashwal, L. D. *Greenstone Belts* (Oxford Scientific, 1997).
9. Kump, L. R. & Barley, M. E. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* **448,** 1033–1036 (2007).

doi:10.1038/nature11275

ASTRONOMY

Planets on the spot

The three planets of the Kepler-30 system align closely with a starspot, indicating their common birth in a gaseous disk. Similar alignments could inform us about the origin of planets orbiting our stellar neighbours. SEE LETTER P.449

DRAKE DEMING

The structure of the Solar System, with multiple planets orbiting in nearly the same geometric plane, is familiar to most. The standard model for the formation of the Solar System hypothesizes that the Sun and planets arose from a spinning disk of gas. Planetary formation in a flat disk naturally produces orbits that are closely aligned in the same geometric plane. But we do not yet know whether the Solar System is typical of planetary systems in general. In exploring that, NASA's Kepler mission is transforming our understanding of planetary formation and evolution. Kepler targets a special category of planets — those that transit, or pass in front of, their star. On page 449 of this issue, Sanchis-Ojeda and collaborators report[1] the fascinating case of the planetary system dubbed Kepler-30. The researchers deduce that all three transiting planets in this system orbit in the same plane, to within one angular degree of inclination. This very tight arrangement in Kepler-30 immediately brings to mind the closely aligned orbits of the Solar System.

One of the first and most spectacular results from Kepler was the discovery that some stars host multiple transiting planets, six being the current record[2]. The mere fact that more than one planet is seen to transit a given star usually implies that they orbit in almost the same geometric plane. The maximum probable difference in inclination for multiple transiting planets is strongly related to the size of their host star, and to the distance of their orbits from the star. The smaller the star and the larger the orbits of the transiting planets, the more likely it is that they lie in the same geometric plane — otherwise we would probably not be able to detect them all transiting their star. One rare exception is where the mutual orbital inclinations of the planets are large, but their orbital planes intersect in a line of nodes that happens — with a minuscule probability — to point towards Earth.

Sanchis-Ojeda *et al.* have taken this trick of geometry a step further. The host star in the Kepler-30 system is magnetically active and covered by starspots — regions that are darker and cooler that the rest of the star's surface. The authors find that the star's three transiting planets all transit the same starspot (Fig. 1), and do so repeatedly. Because the starspot is much smaller than the star itself, this regularity requires the planetary orbits to be aligned very closely, and also to be aligned with the rotational equator of the star. Like the Solar System, the planets of Kepler-30 clearly originated in a spinning disk of gas.

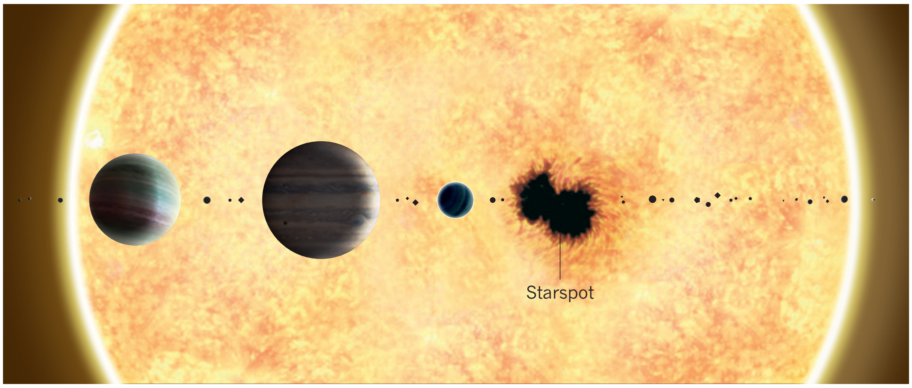


Figure 1 | Planetary line-up. Sanchis-Ojeda *et al.*[1] find that the three planets of the Kepler-30 system, seen here in an artist's impression, all transit the same starspot, indicating that they lie in the same geometric plane. The plane in which they orbit may well contain moonlets, asteroids and other circumstellar components.

Lest we assume that all extrasolar planets remain aligned in disks, Sanchis-Ojeda *et al.* point out the puzzling orbital properties of transiting giant planets that orbit close to their parent stars. Many of these 'hot Jupiters' are strongly misaligned with the equators of their stars[3]. The fact that some planetary orbits can be neatly aligned with the plane of their star's equator, whereas others are wildly misaligned, shows that we have much still to learn about the formation and orbital evolution of planets.

Finding planets that line up with starspots might seem like a lucky one-off. But the history of extrasolar planetary science shows that today's one-time occurrence is tomorrow's standard technique. The measurement of exoplanetary transmission spectra is a prime example. The detection of sodium absorption in the atmosphere of a giant transiting planet by the Hubble Space Telescope[4] seemed at first to be a singular observation, difficult to apply to other worlds. Now, the technique of transmission spectroscopy is widely applied[5] and projected to be a mainstay of future exoplanetary spectroscopy[6].

There is reason to believe that studying the alignment of transiting planets with starspots could be increasingly important in the future. Currently, NASA is contemplating the Transiting Exoplanet Survey Satellite (TESS), to survey the entire sky for planets transiting nearby stars. An all-sky survey such as TESS preferentially samples the volume of space in our local solar neighbourhood, where the overwhelming majority of stars are small, cool 'M-dwarfs' and 'K-dwarfs'. The interiors of these small red stars are constantly stirred by convective motions, generating strong magnetic fields that thread the stars' outer layers and produce starspots. Moreover, we know that small stars commonly host a class of rocky planets called 'super-Earths'[7]. With masses up to 20 times that of Earth, and radii up to 3 times the radius of Earth, super-Earths form in abundance around small, low-mass stars. Presumably the formation of super-Earths occurs in protoplanetary disks of gas and dust[8], but we know precious little about planetary formation around low-mass stars.

Protoplanetary disks surrounding small stars are of low mass and brightness, and are difficult to observe. Yet it is not too far-fetched to imagine that we may be able to infer much about the formation of planets in our immediate solar neighbourhood by observing the transiting planets that align with starspots, as in Kepler-30. Today's line-up of 'planets on the spot' is a glimpse into the future of exoplanetary science. ■

Drake Deming *is in the Department of Astronomy, University of Maryland at College Park, College Park, Maryland 20742, USA.*
e-mail: ddeming@astro.umd.edu

1. Sanchis-Ojeda, R. *et al. Nature* **487,** 449–453 (2012).
2. Lissauer, J. J. *et al. Nature* **470,** 53–58 (2011).
3. Winn, J. N., Fabrycky, D., Albrecht, S. & Johnson, J. A. *Astrophys. J.* **718,** L145–L149 (2010).
4. Charbonneau, D. *et al. Astrophys. J.* **568,** 377–384 (2002).
5. Berta, Z. K. *et al. Astrophys. J.* **747,** 35 (2012).
6. Deming, D. *et al. Publ. Astron. Soc. Pacif.* **121,** 952–967 (2009).
7. Howard, A. W. *et al. Astrophys. J. Suppl.* **201,** 15 (2012).
8. Montgomery, R. & Laughlin, G. *Icarus* **202,** 1–11 (2009).

NEUROSCIENCE

The wrap that feeds neurons

The myelin sheath around nerve fibres serves to speed up electrical nerve signals. But it turns out that it also supplies neurons with fuel to support their high metabolic activity. See Article p.443

JOHANNE E. RINHOLM & LINDA H. BERGERSEN

The white matter of the central nervous system consists of nerve fibres (axons) wrapped in insulating myelin — a membranous, lipid-rich sheath that is laid down by cells known as oligodendrocytes. In the peripheral nervous system, fast-conducting nerve fibres are similarly wrapped in myelin by Schwann cells. The sheath allows nerve impulses to move at a much higher speed than in non-myelinated axons, and supports neurons in other, less-well-understood ways[1]. On page 443 of this issue, Lee *et al.*[2] describe how the myelin wrap provides lactate, an energy-generating metabolite, to axons.

Cells harness metabolic energy by producing the molecule ATP from glucose. Some ATP is generated in the cytoplasm by the breakdown of glucose to pyruvate and lactate, in a process known as glycolysis. But most of the metabolic energy is obtained by the oxidation of pyruvate and lactate in mitochondria, the cell's powerhouses. Neurons can oxidize lactate that is released from astrocytes, a type of supporting cell that can sustain high rates of glycolysis (Fig. 1). The release of lactate from astrocytes is vital for neuronal functions[3] such as memory formation[4], as well as for neuronal survival under several forms of stress[5]. Lactate also provides carbon for the synthesis of cell constituents such as lipids[6], and might signal a cell's metabolic state by binding to specific receptor proteins on the surface of neighbouring cells[7].

In the brain, lactate is transferred between cells through transporter proteins known as MCT1, MCT2 and MCT4. Previous research suggested that MCT1 is expressed mainly in endothelial cells (which form the walls of blood vessels), MCT2 in neurons and MCT4 in astrocytes[8]. In addition, it has been shown[6,9] that oligodendrocytes use MCT1 to take up lactate, which helps them to survive and produce myelin, particularly in low-glucose conditions. However, once myelination has taken place, oligodendrocytes seem to rely mainly on non-mitochondrial metabolic pathways — such as glycolysis — for energy production[10]. As a result, oligodendrocytes release lactate that might be used by active neurons.

Lee *et al.* found MCT1 expression in oligodendrocytes but — in contrast to some findings[8,9] — not in astrocytes or endothelial cells. The researchers induced a reduction of *MCT1* gene expression in cultures of spinal cord tissue and in live mice, and observed the formation of abnormal axons and neuronal death. What's more, when they combined such *MCT1* knockdown with glucose deprivation in spinal cord cultures, the tissue suffered extensive neuronal death, which could be prevented by supplying exogenous lactate. Therefore, the *MCT1* knockdown blocked lactate export from oligodendrocytes, but not its import into axons. These results indicate that the myelin sheath does not physically block the access of axons to exogenous metabolites, and that neurons need the lactate provided by oligodendrocytes when that released to the extracellular milieu by other cells such as astrocytes (Fig. 1) is insufficient.

The authors noticed that the changes caused in mice by *MCT1* knockdown were similar to those seen in patients with amyotrophic lateral sclerosis (ALS) — a fatal neurodegenerative disease that is characterized by progressive loss of spinal cord motor neurons, which control muscle movement. They found that MCT1 expression in tissue from patients with ALS, and from genetically engineered mice that are commonly used as a model of the disease, was lower than that in tissues from healthy people and wild-type mice. This suggests that reduced MCT1 activity could be involved in ALS disease progression. However, the expression of MCT4 was also reduced in patients' tissue, presumably resulting in reduced

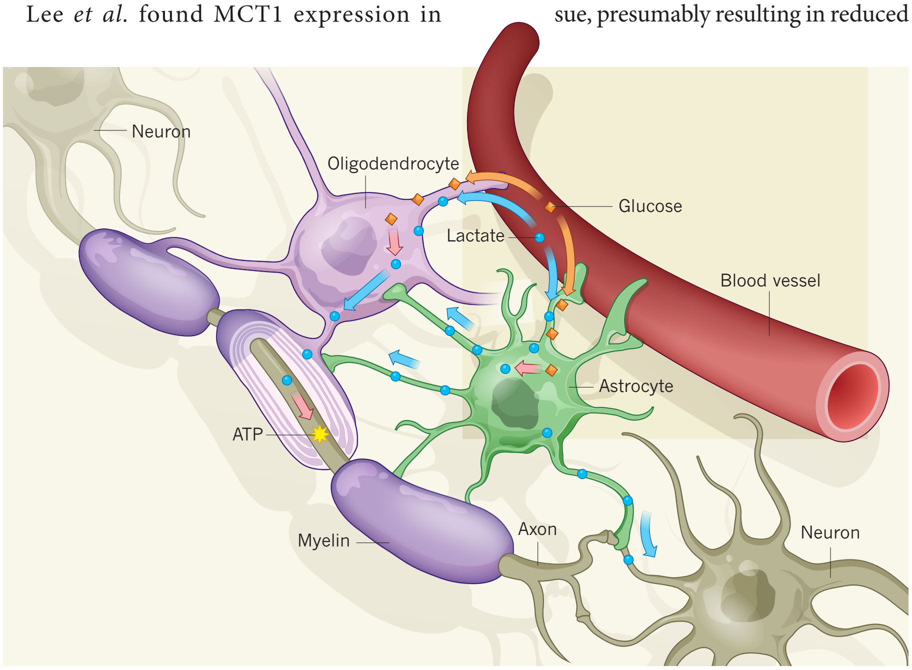


Figure 1 | Metabolic support for neurons. Blood glucose (orange) can be taken up by non-neuronal brain cells known as oligodendrocytes and astrocytes. Inside the cells, glucose is broken down to lactate (blue), which can then be transferred to neurons and used to generate metabolic energy in the form of ATP molecules (yellow). Lee *et al.*[2] show that oligodendrocytes supply nerve fibres (axons) with lactate delivered through a transporter protein (MCT1, not shown) located in the myelin, a membranous sheath around the axons. Lactate can also be produced in astrocytes and then transferred to axons by means of small pores (gap junctions, not shown) through the myelin. In addition, blood lactate may reach neurons through astrocytes and oligodendrocytes.

Drake Deming *is in the Department of Astronomy, University of Maryland at College Park, College Park, Maryland 20742, USA.*
e-mail: ddeming@astro.umd.edu

1. Sanchis-Ojeda, R. *et al. Nature* **487,** 449–453 (2012).
2. Lissauer, J. J. *et al. Nature* **470,** 53–58 (2011).
3. Winn, J. N., Fabrycky, D., Albrecht, S. & Johnson, J. A. *Astrophys. J.* **718,** L145–L149 (2010).
4. Charbonneau, D. *et al. Astrophys. J.* **568,** 377–384 (2002).
5. Berta, Z. K. *et al. Astrophys. J.* **747,** 35 (2012).
6. Deming, D. *et al. Publ. Astron. Soc. Pacif.* **121,** 952–967 (2009).
7. Howard, A. W. *et al. Astrophys. J. Suppl.* **201,** 15 (2012).
8. Montgomery, R. & Laughlin, G. *Icarus* **202,** 1–11 (2009).

NEUROSCIENCE

The wrap that feeds neurons

The myelin sheath around nerve fibres serves to speed up electrical nerve signals. But it turns out that it also supplies neurons with fuel to support their high metabolic activity. See Article p.443

JOHANNE E. RINHOLM & LINDA H. BERGERSEN

The white matter of the central nervous system consists of nerve fibres (axons) wrapped in insulating myelin — a membranous, lipid-rich sheath that is laid down by cells known as oligodendrocytes. In the peripheral nervous system, fast-conducting nerve fibres are similarly wrapped in myelin by Schwann cells. The sheath allows nerve impulses to move at a much higher speed than in non-myelinated axons, and supports neurons in other, less-well-understood ways[1]. On page 443 of this issue, Lee *et al.*[2] describe how the myelin wrap provides lactate, an energy-generating metabolite, to axons.

Cells harness metabolic energy by producing the molecule ATP from glucose. Some ATP is generated in the cytoplasm by the breakdown of glucose to pyruvate and lactate, in a process known as glycolysis. But most of the metabolic energy is obtained by the oxidation of pyruvate and lactate in mitochondria, the cell's powerhouses. Neurons can oxidize lactate that is released from astrocytes, a type of supporting cell that can sustain high rates of glycolysis (Fig. 1). The release of lactate from astrocytes is vital for neuronal functions[3] such as memory formation[4], as well as for neuronal survival under several forms of stress[5]. Lactate also provides carbon for the synthesis of cell constituents such as lipids[6], and might signal a cell's metabolic state by binding to specific receptor proteins on the surface of neighbouring cells[7].

In the brain, lactate is transferred between cells through transporter proteins known as MCT1, MCT2 and MCT4. Previous research suggested that MCT1 is expressed mainly in endothelial cells (which form the walls of blood vessels), MCT2 in neurons and MCT4 in astrocytes[8]. In addition, it has been shown[6,9] that oligodendrocytes use MCT1 to take up lactate, which helps them to survive and produce myelin, particularly in low-glucose conditions. However, once myelination has taken place, oligodendrocytes seem to rely mainly on non-mitochondrial metabolic pathways — such as glycolysis — for energy production[10]. As a result, oligodendrocytes release lactate that might be used by active neurons.

Lee *et al.* found MCT1 expression in oligodendrocytes but — in contrast to some findings[8,9] — not in astrocytes or endothelial cells. The researchers induced a reduction of *MCT1* gene expression in cultures of spinal cord tissue and in live mice, and observed the formation of abnormal axons and neuronal death. What's more, when they combined such *MCT1* knockdown with glucose deprivation in spinal cord cultures, the tissue suffered extensive neuronal death, which could be prevented by supplying exogenous lactate. Therefore, the *MCT1* knockdown blocked lactate export from oligodendrocytes, but not its import into axons. These results indicate that the myelin sheath does not physically block the access of axons to exogenous metabolites, and that neurons need the lactate provided by oligodendrocytes when that released to the extracellular milieu by other cells such as astrocytes (Fig. 1) is insufficient.

The authors noticed that the changes caused in mice by *MCT1* knockdown were similar to those seen in patients with amyotrophic lateral sclerosis (ALS) — a fatal neurodegenerative disease that is characterized by progressive loss of spinal cord motor neurons, which control muscle movement. They found that MCT1 expression in tissue from patients with ALS, and from genetically engineered mice that are commonly used as a model of the disease, was lower than that in tissues from healthy people and wild-type mice. This suggests that reduced MCT1 activity could be involved in ALS disease progression. However, the expression of MCT4 was also reduced in patients' tissue, presumably resulting in reduced

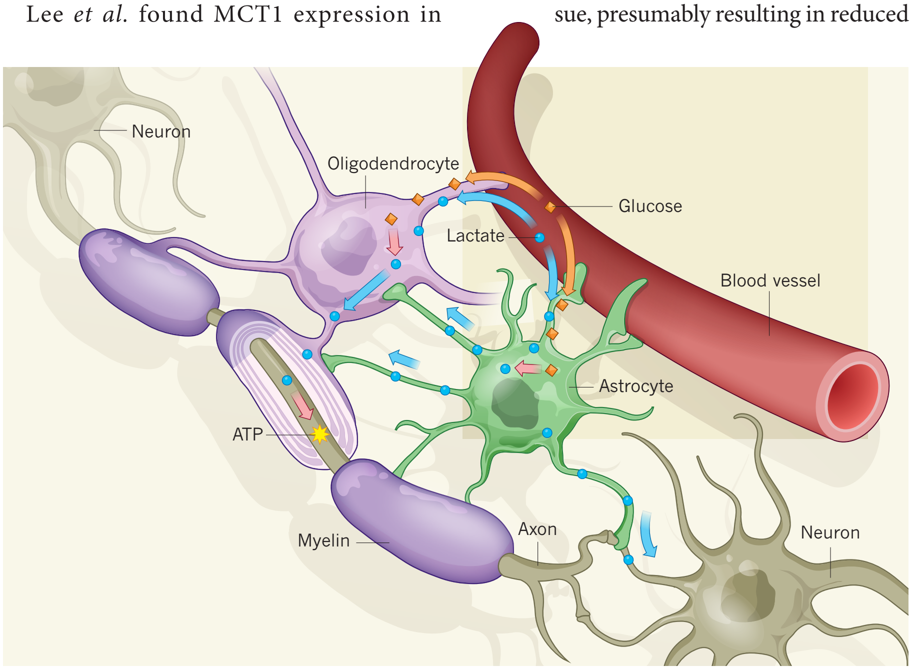


Figure 1 | Metabolic support for neurons. Blood glucose (orange) can be taken up by non-neuronal brain cells known as oligodendrocytes and astrocytes. Inside the cells, glucose is broken down to lactate (blue), which can then be transferred to neurons and used to generate metabolic energy in the form of ATP molecules (yellow). Lee *et al.*[2] show that oligodendrocytes supply nerve fibres (axons) with lactate delivered through a transporter protein (MCT1, not shown) located in the myelin, a membranous sheath around the axons. Lactate can also be produced in astrocytes and then transferred to axons by means of small pores (gap junctions, not shown) through the myelin. In addition, blood lactate may reach neurons through astrocytes and oligodendrocytes.

lactate supply and hence contributing to neuronal energy failure. The expression of MCT2 was not tested by the authors.

Lee and colleagues' results, together with those of other researchers[10], show that lactate provided by or through the myelin sheath is crucial for neuronal survival. Such lactate could be produced directly in oligodendrocytes by glycolysis, or could be transferred to these cells from astrocytes through gap junctions (small pores that connect the cytoplasm of two adjacent cells). In addition, both oligodendrocytes and astrocytes could take up lactate from the blood (Fig. 1).

Notably, the authors found that lactate from oligodendrocytes was particularly important for the survival of motor neurons. This could be because motor neurons have very long axons, and so the supply of lactate or pyruvate from the distant cell body (where glycolysis mainly takes place) might be less efficient than in neurons with shorter axons.

As Lee *et al.* point out, insufficient neuronal energy production has already been implicated in ALS. It is also known[11] that a mutation in the gene encoding superoxide dismutase 1, an enzyme involved in the removal of potentially harmful reactive oxygen species, leads to mitochondrial dysfunction and causes a hereditary form of ALS. The reduced MCT1 expression that the authors observed in patients with ALS might be linked to mitochondrial dysfunction, because a lack of energy-providing substrates (such as lactate) could damage the mitochondria. It would therefore be interesting to test whether overexpression of MCT1 prevents disease progression in the ALS mouse model. Furthermore, the researchers' findings are relevant to other neurological disorders, because neuronal energy deficiency has been implicated in epilepsy and demyelinating disorders, as well as in Alzheimer's, Parkinson's and Huntington's diseases[12].

The paper by Lee *et al.* expands the known roles of myelin sheaths and reveals how their dysfunction might cause disease. Nevertheless, to solve the mysteries of ALS and other neurological disorders, we probably need to broaden the focus of our research from neurons to other brain cells. In addition, the authors' demonstration that exogenously supplied lactate can overcome the neuronal dysfunction caused by deficient lactate supply from oligodendrocytes is intriguing. This suggests that lactate has a therapeutic potential, and that one of the ways in which physical exercise can benefit the brain is by increasing lactate supply through the bloodstream. ■

Johanne E. Rinholm *and* **Linda H. Bergersen** *are at the Centre for Molecular Biology and Neuroscience, and the Department of Anatomy, Institute of Basic Medical Sciences, University of Oslo, Oslo N-0317, Norway.* **J.E.R.** *is also at the Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA.* **L.H.B.** *is also at the Center for Healthy Aging, Department of Neuroscience and Pharmacology, University of Copenhagen, 2200 Copenhagen N, Denmark.*
e-mails: j.e.rinholm@medisin.uio.no; l.h.bergersen@medisin.uio.no

1. Nave, K.-A. *Nature Rev. Neurosci.* **11,** 275–283 (2010).
2. Lee, Y. *et al. Nature* **487,** 443–448 (2012).
3. Pellerin, L. & Magistretti, P. J. *J. Cereb. Blood Flow Metab.* **32,** 1152–1166 (2012).
4. Suzuki, A. *et al. Cell* **144,** 810–823 (2011).
5. Won, S. J. *et al. J. Cereb. Blood Flow Metab.* **32,** 1086–1096 (2012).
6. Sánchez-Abarca, L. I., Tabernero, A. & Medina, J. M. *Glia* **36,** 321–329 (2001).
7. Bergersen, L. H. & Gjedde, A. *Front. Neuroenerg.* **4,** 5 (2012).
8. Bergersen, L. H. *Neuroscience* **145,** 11–19 (2007).
9. Rinholm, J. E. *et al. J. Neurosci.* **31,** 538–548 (2011).
10. Fünfschilling, U. *et al. Nature* **485,** 517–521 (2012).
11. Carrì, M. T. & Cozzolino, M. *J. Bioenerg. Biomembr.* **43,** 593–599 (2011).
12. Lax, N. Z., Turnbull, D. M. & Reeve, A. K. *Neuroscientist* **17,** 645–658 (2011).

DEVICE PHYSICS

Put the pedal to the metal

The discovery of a powerful way to switch a material from an insulator to a metal might ultimately be exploited to build a new generation of electronic switches.
SEE LETTER P.459

JOCHEN MANNHART & WILFRIED HAENSCH

What a concept: to use an electrical voltage to turn an insulator into a metal and back, super fast and almost without energy losses. This idea of an 'ultimate electronic switch' goes back to the early twentieth century, when researchers tried to change the electrical resistance of materials by adding or removing mobile charges. These charges are generated by an electric field arising from a voltage applied to an electrode, known as the gate, that is laid on top of the materials. As early as 1925, a patent was filed[1] for devices that change the resistance of a conducting material. These devices are now called field-effect transistors. On page 459 of this issue, Nakano *et al.*[2] report a type of field-effect device that may pave the way to the ultimate switch.

In a field-effect transistor (FET), one elementary charge put onto the gate enhances the mobile charge-carrier density in the conducting channel of the transistor by exactly one; less, if some of the charge carriers induced in the channel get trapped at material defects. Such electric-field switching of semiconductor channels forms the basis of almost all electronic devices in use today. It is the veritable backbone of our electronic society. The switching is also useful for fundamental science, because it allows us to study materials by varying the electron density in a given chunk of matter[3].

Until now, the crux of switching has been that it occurs only in a thin surface layer rather than in the bulk of the channel material — not unlike a sports car's accelerator being blocked after being pushed down by only a fraction. What's more, with a few exceptions, the electric field modulates only the resistance of semiconductors — it does not turn metals into insulators and vice versa, which could enable improved transistors to be made.

Enter Nakano and his collaborators. In their experiments, the authors used huge electric fields generated by a voltage applied to a drop of an ionic liquid (a fluid composed mostly of ions) on a film of vanadium dioxide (VO_2). VO_2 is a one-of-a-kind solid: at room temperature, it is an insulator, but, if warmed to 340 kelvin, it turns into a metal. Furthermore, the electrons in VO_2 are correlated — the precise behaviour of an electron depends on what its neighbours are doing.

Nakano and colleagues' experiments revealed that applying a voltage to the ionic liquid switches VO_2 from an insulator to a metal. The big surprise was that this transition occurs not only in a thin layer of the material, but over the complete thickness of all of the samples studied — with 70 nanometres the maximum thickness tested. The occurrence of the transition over the complete thickness compares to 'putting the pedal to the metal' in a sports car. Furthermore, in contrast to the conventional FET, when VO_2 is switched from an insulator to a metal many more mobile charge carriers are generated in the VO_2 than are put onto the gate, because many pre-existing immobile charge carriers are freed and delocalized by this transition (Fig. 1). All of this occurs at close to room temperature.

For experiments such as these, however, pitfalls lurk everywhere. As a result, the authors used three independent techniques — X-ray diffraction, resistance and 'Hall-effect'

lactate supply and hence contributing to neuronal energy failure. The expression of MCT2 was not tested by the authors.

Lee and colleagues' results, together with those of other researchers[10], show that lactate provided by or through the myelin sheath is crucial for neuronal survival. Such lactate could be produced directly in oligodendrocytes by glycolysis, or could be transferred to these cells from astrocytes through gap junctions (small pores that connect the cytoplasm of two adjacent cells). In addition, both oligodendrocytes and astrocytes could take up lactate from the blood (Fig. 1).

Notably, the authors found that lactate from oligodendrocytes was particularly important for the survival of motor neurons. This could be because motor neurons have very long axons, and so the supply of lactate or pyruvate from the distant cell body (where glycolysis mainly takes place) might be less efficient than in neurons with shorter axons.

As Lee *et al.* point out, insufficient neuronal energy production has already been implicated in ALS. It is also known[11] that a mutation in the gene encoding superoxide dismutase 1, an enzyme involved in the removal of potentially harmful reactive oxygen species, leads to mitochondrial dysfunction and causes a hereditary form of ALS. The reduced MCT1 expression that the authors observed in patients with ALS might be linked to mitochondrial dysfunction, because a lack of energy-providing substrates (such as lactate) could damage the mitochondria. It would therefore be interesting to test whether overexpression of MCT1 prevents disease progression in the ALS mouse model. Furthermore, the researchers' findings are relevant to other neurological disorders, because neuronal energy deficiency has been implicated in epilepsy and demyelinating disorders, as well as in Alzheimer's, Parkinson's and Huntington's diseases[12].

The paper by Lee *et al.* expands the known roles of myelin sheaths and reveals how their dysfunction might cause disease. Nevertheless, to solve the mysteries of ALS and other neurological disorders, we probably need to broaden the focus of our research from neurons to other brain cells. In addition, the authors' demonstration that exogenously supplied lactate can overcome the neuronal dysfunction caused by deficient lactate supply from oligodendrocytes is intriguing. This suggests that lactate has a therapeutic potential, and that one of the ways in which physical exercise can benefit the brain is by increasing lactate supply through the bloodstream. ■

Johanne E. Rinholm *and* **Linda H. Bergersen** *are at the Centre for Molecular Biology and Neuroscience, and the Department of Anatomy, Institute of Basic Medical Sciences, University of Oslo, Oslo N-0317, Norway.* **J.E.R.** *is also at the Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA.* **L.H.B.** *is also at the Center for Healthy Aging, Department of Neuroscience and Pharmacology, University of Copenhagen, 2200 Copenhagen N, Denmark.*
e-mails: j.e.rinholm@medisin.uio.no; l.h.bergersen@medisin.uio.no

1. Nave, K.-A. *Nature Rev. Neurosci.* **11,** 275–283 (2010).
2. Lee, Y. *et al. Nature* **487,** 443–448 (2012).
3. Pellerin, L. & Magistretti, P. J. *J. Cereb. Blood Flow Metab.* **32,** 1152–1166 (2012).
4. Suzuki, A. *et al. Cell* **144,** 810–823 (2011).
5. Won, S. J. *et al. J. Cereb. Blood Flow Metab.* **32,** 1086–1096 (2012).
6. Sánchez-Abarca, L. I., Tabernero, A. & Medina, J. M. *Glia* **36,** 321–329 (2001).
7. Bergersen, L. H. & Gjedde, A. *Front. Neuroenerg.* **4,** 5 (2012).
8. Bergersen, L. H. *Neuroscience* **145,** 11–19 (2007).
9. Rinholm, J. E. *et al. J. Neurosci.* **31,** 538–548 (2011).
10. Fünfschilling, U. *et al. Nature* **485,** 517–521 (2012).
11. Carrì, M. T. & Cozzolino, M. *J. Bioenerg. Biomembr.* **43,** 593–599 (2011).
12. Lax, N. Z., Turnbull, D. M. & Reeve, A. K. *Neuroscientist* **17,** 645–658 (2011).

DEVICE PHYSICS

Put the pedal to the metal

The discovery of a powerful way to switch a material from an insulator to a metal might ultimately be exploited to build a new generation of electronic switches.

SEE LETTER P.459

JOCHEN MANNHART & WILFRIED HAENSCH

What a concept: to use an electrical voltage to turn an insulator into a metal and back, super fast and almost without energy losses. This idea of an 'ultimate electronic switch' goes back to the early twentieth century, when researchers tried to change the electrical resistance of materials by adding or removing mobile charges. These charges are generated by an electric field arising from a voltage applied to an electrode, known as the gate, that is laid on top of the materials. As early as 1925, a patent was filed[1] for devices that change the resistance of a conducting material. These devices are now called field-effect transistors. On page 459 of this issue, Nakano *et al.*[2] report a type of field-effect device that may pave the way to the ultimate switch.

In a field-effect transistor (FET), one elementary charge put onto the gate enhances the mobile charge-carrier density in the conducting channel of the transistor by exactly one; less, if some of the charge carriers induced in the channel get trapped at material defects. Such electric-field switching of semiconductor channels forms the basis of almost all electronic devices in use today. It is the veritable backbone of our electronic society. The switching is also useful for fundamental science, because it allows us to study materials by varying the electron density in a given chunk of matter[3].

Until now, the crux of switching has been that it occurs only in a thin surface layer rather than in the bulk of the channel material — not unlike a sports car's accelerator being blocked after being pushed down by only a fraction. What's more, with a few exceptions, the electric field modulates only the resistance of semiconductors — it does not turn metals into insulators and vice versa, which could enable improved transistors to be made.

Enter Nakano and his collaborators. In their experiments, the authors used huge electric fields generated by a voltage applied to a drop of an ionic liquid (a fluid composed mostly of ions) on a film of vanadium dioxide (VO_2). VO_2 is a one-of-a-kind solid: at room temperature, it is an insulator, but, if warmed to 340 kelvin, it turns into a metal. Furthermore, the electrons in VO_2 are correlated — the precise behaviour of an electron depends on what its neighbours are doing.

Nakano and colleagues' experiments revealed that applying a voltage to the ionic liquid switches VO_2 from an insulator to a metal. The big surprise was that this transition occurs not only in a thin layer of the material, but over the complete thickness of all of the samples studied — with 70 nanometres the maximum thickness tested. The occurrence of the transition over the complete thickness compares to 'putting the pedal to the metal' in a sports car. Furthermore, in contrast to the conventional FET, when VO_2 is switched from an insulator to a metal many more mobile charge carriers are generated in the VO_2 than are put onto the gate, because many pre-existing immobile charge carriers are freed and delocalized by this transition (Fig. 1). All of this occurs at close to room temperature.

For experiments such as these, however, pitfalls lurk everywhere. As a result, the authors used three independent techniques — X-ray diffraction, resistance and 'Hall-effect'

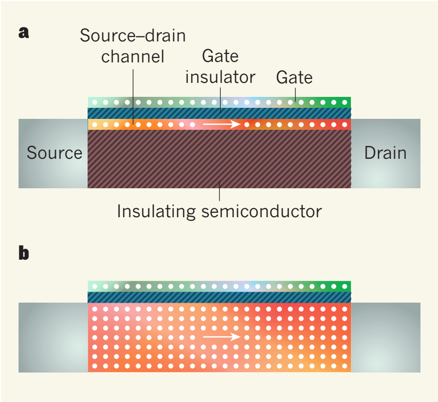


Figure 1 | Cross-sectional view of the 'on state' of two types of transistor. a, When an electrical voltage is applied to a conventional field-effect transistor, electrons (white dots) added to the device's gate electrode generate mobile electrons. These flow (arrow) from the source to the drain electrodes in a channel that consists of a thin, conducting layer of semiconductor material. This channel is sandwiched between the gate insulator and the insulating part of the semiconductor. **b**, In a transistor that is based on a material that undergoes an insulator–metal phase transition, such as the transition observed by Nakano *et al.*[2] in vanadium dioxide, the channel through which the electrons flow extends across the complete thickness of the material. The number of mobile electrons in the channel greatly exceeds the number of electrons put onto the gate.

measurements — to investigate whether the switching occurs across the entire thickness of the VO_2. All of the methods showed that this is the case. It seems that the ionic liquid starts by switching a surface layer, thereby inducing an insulator–metal phase boundary in the VO_2. This phase boundary seems to be so energy intensive that the insulator–metal phase transition rattles through the rest of the film, driven by collective deformations of the material's lattice structure. To what extent the switching is triggered by the electric field and/or by the charged molecules of the ionic liquid interacting with the sample surface remains an open question.

These experiments show that voltage-controlled insulator–metal transitions are possible and can result in sizable changes in the electronic properties of the material. The delocalization of immobile bulk charges in the VO_2 films owing to the application of an electric field to their surface is a phenomenon that might ultimately be exploited to build a new generation of electronic switches. The mobile charge-carrier density that can be attained in such channels is significantly larger than that in conventional FETs, and high carrier densities are sought after for further miniaturization of transistors.

The authors observed the insulator–metal transition, and so the switch from an 'on' to an 'off' state, in a voltage and temperature range that is realistic for applications. The VO_2 device displayed a tenfold increase in the channel current for a 100-millivolt change in the applied gate voltage — to use the jargon, an on/off swing of about 100 mV per decade. The measured ratio of the on-current to the off-current was about 100, and the device could, in principle, operate at a gate voltage of 1 V. For comparison, in today's low-power handheld electronic devices, an on/off current ratio of one million is desirable, with an on/off swing of approximately 90 mV per decade at a gate voltage of about 1.5 V.

The possibility of operating the device at a gate voltage of 1 V offers the prospect of using it in low-power technology, provided that the switching behaviour can be improved. Structural changes that have a role in the switching often happen at a slower rate than electronic changes, and so might limit the switching speed. However, even if the phase transition propagates through the film's thickness only at the speed of sound, switching times far below one nanosecond may eventually be possible. The burning question for possible applications is whether the switching can also be achieved in devices in which the ionic liquid is replaced with a solid insulator, as required for the integration of a device into an electronic circuit. And, of course, it would be interesting to know the maximum device thickness for which switching can be attained. Further insight into the switching mechanism of the device, and progress in material engineering, might lead the way towards such a solid-state switch.

Because the switching may occur by means of a phase transition rattling at high speed through matter, possible device applications could extend far beyond FETs as we know them. The number of candidate materials that could be used to achieve similar switching is enormous and, as the predictive power of *ab initio* calculations continues to advance, chances are that material combinations can be found that will improve the switching behaviour of the device, making it suitable for use in electronic circuits. ■

Jochen Mannhart *is at the Max Planck Institute for Solid State Research, 70569 Stuttgart, Germany.* **Wilfried Haensch** *is at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA.*
e-mails: j.mannhart@fkf.mpg.de; whaensch@us.ibm.com

1. Lilienfeld, J. E. US Patent 1745175 (1930).
2. Nakano, M. *et al. Nature* **487,** 459–462 (2012).
3. Ahn, C. H., Triscone, J.-M. & Mannhart, J. *Nature* **424,** 1015–1018 (2003).

IMMUNOLOGY

Malnutrition promotes rogue bacteria

Dietary lack of a single amino acid impairs intestinal immunity in mice, altering the gut's microbial community and leaving it vulnerable to damage. The finding helps to explain how malnutrition favours gut inflammation. **SEE LETTER P.477**

ANA IZCUE & FIONA POWRIE

Malnutrition enhances the susceptibility to and the severity of gastrointestinal infections and diarrhoeal diseases[1]. Although this effect has been observed for decades, the underlying mechanisms linking diet and immunodeficiency have remained largely unknown. On page 477 of this issue, Hashimoto *et al.*[2] elegantly dissect the connection between protein deficiency and intestinal inflammation, uncovering an intricate network involving nutrient transport, microbial ecology, antimicrobial responses and inflammation.

Most amino acids can be synthesized in the body from other compounds, but the so-called essential amino acids must be obtained from the diet through the action of specialized amino-acid transport proteins. The starting point of Hashimoto and colleagues' studies were mice lacking the gene encoding angiotensin I converting enzyme 2 (ACE2). ACE2 controls the functional expression in the intestines of one of the transport proteins, B^0AT1, which acts specifically on neutral amino acids. The authors found that the ACE2-deficient mice had low levels of neutral essential amino acids in their blood, but no visible alterations in the architecture of their gut tissues. But the findings became more interesting when the authors exposed the mutant mice to a chemical irritant, to test their sensitivity to intestinal inflammation. Although control mice showed only moderate inflammatory symptoms following this chemical challenge, the ACE2-deficient mice were much more vulnerable, showing increased diarrhoea and greater numbers of inflammatory cells. This result was surprising, because this irritation procedure induces inflammation in the large intestine, a part of the gut in which ACE2 is not highly expressed.

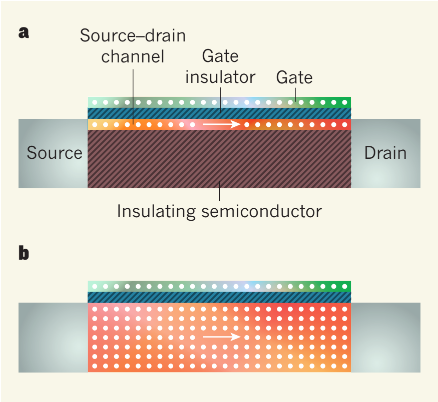


Figure 1 | Cross-sectional view of the 'on state' of two types of transistor. a, When an electrical voltage is applied to a conventional field-effect transistor, electrons (white dots) added to the device's gate electrode generate mobile electrons. These flow (arrow) from the source to the drain electrodes in a channel that consists of a thin, conducting layer of semiconductor material. This channel is sandwiched between the gate insulator and the insulating part of the semiconductor. **b**, In a transistor that is based on a material that undergoes an insulator–metal phase transition, such as the transition observed by Nakano *et al.*[2] in vanadium dioxide, the channel through which the electrons flow extends across the complete thickness of the material. The number of mobile electrons in the channel greatly exceeds the number of electrons put onto the gate.

measurements — to investigate whether the switching occurs across the entire thickness of the VO_2. All of the methods showed that this is the case. It seems that the ionic liquid starts by switching a surface layer, thereby inducing an insulator–metal phase boundary in the VO_2. This phase boundary seems to be so energy intensive that the insulator–metal phase transition rattles through the rest of the film, driven by collective deformations of the material's lattice structure. To what extent the switching is triggered by the electric field and/or by the charged molecules of the ionic liquid interacting with the sample surface remains an open question.

These experiments show that voltage-controlled insulator–metal transitions are possible and can result in sizable changes in the electronic properties of the material. The delocalization of immobile bulk charges in the VO_2 films owing to the application of an electric field to their surface is a phenomenon that might ultimately be exploited to build a new generation of electronic switches. The mobile charge-carrier density that can be attained in such channels is significantly larger than that in conventional FETs, and high carrier densities are sought after for further miniaturization of transistors.

The authors observed the insulator–metal transition, and so the switch from an 'on' to an 'off' state, in a voltage and temperature range that is realistic for applications. The VO_2 device displayed a tenfold increase in the channel current for a 100-millivolt change in the applied gate voltage — to use the jargon, an on/off swing of about 100 mV per decade. The measured ratio of the on-current to the off-current was about 100, and the device could, in principle, operate at a gate voltage of 1 V. For comparison, in today's low-power handheld electronic devices, an on/off current ratio of one million is desirable, with an on/off swing of approximately 90 mV per decade at a gate voltage of about 1.5 V.

The possibility of operating the device at a gate voltage of 1 V offers the prospect of using it in low-power technology, provided that the switching behaviour can be improved. Structural changes that have a role in the switching often happen at a slower rate than electronic changes, and so might limit the switching speed. However, even if the phase transition propagates through the film's thickness only at the speed of sound, switching times far below one nanosecond may eventually be possible. The burning question for possible applications is whether the switching can also be achieved in devices in which the ionic liquid is replaced with a solid insulator, as required for the integration of a device into an electronic circuit. And, of course, it would be interesting to know the maximum device thickness for which switching can be attained. Further insight into the switching mechanism of the device, and progress in material engineering, might lead the way towards such a solid-state switch.

Because the switching may occur by means of a phase transition rattling at high speed through matter, possible device applications could extend far beyond FETs as we know them. The number of candidate materials that could be used to achieve similar switching is enormous and, as the predictive power of *ab initio* calculations continues to advance, chances are that material combinations can be found that will improve the switching behaviour of the device, making it suitable for use in electronic circuits. ■

Jochen Mannhart *is at the Max Planck Institute for Solid State Research, 70569 Stuttgart, Germany.* **Wilfried Haensch** *is at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA.*
e-mails: j.mannhart@fkf.mpg.de; whaensch@us.ibm.com

1. Lilienfeld, J. E. US Patent 1745175 (1930).
2. Nakano, M. *et al. Nature* **487,** 459–462 (2012).
3. Ahn, C. H., Triscone, J.-M. & Mannhart, J. *Nature* **424,** 1015–1018 (2003).

IMMUNOLOGY

Malnutrition promotes rogue bacteria

Dietary lack of a single amino acid impairs intestinal immunity in mice, altering the gut's microbial community and leaving it vulnerable to damage. The finding helps to explain how malnutrition favours gut inflammation. **SEE LETTER P.477**

ANA IZCUE & FIONA POWRIE

Malnutrition enhances the susceptibility to and the severity of gastrointestinal infections and diarrhoeal diseases[1]. Although this effect has been observed for decades, the underlying mechanisms linking diet and immunodeficiency have remained largely unknown. On page 477 of this issue, Hashimoto *et al.*[2] elegantly dissect the connection between protein deficiency and intestinal inflammation, uncovering an intricate network involving nutrient transport, microbial ecology, antimicrobial responses and inflammation.

Most amino acids can be synthesized in the body from other compounds, but the so-called essential amino acids must be obtained from the diet through the action of specialized amino-acid transport proteins. The starting point of Hashimoto and colleagues' studies were mice lacking the gene encoding angiotensin I converting enzyme 2 (ACE2). ACE2 controls the functional expression in the intestines of one of the transport proteins, B^0AT1, which acts specifically on neutral amino acids. The authors found that the ACE2-deficient mice had low levels of neutral essential amino acids in their blood, but no visible alterations in the architecture of their gut tissues. But the findings became more interesting when the authors exposed the mutant mice to a chemical irritant, to test their sensitivity to intestinal inflammation. Although control mice showed only moderate inflammatory symptoms following this chemical challenge, the ACE2-deficient mice were much more vulnerable, showing increased diarrhoea and greater numbers of inflammatory cells. This result was surprising, because this irritation procedure induces inflammation in the large intestine, a part of the gut in which ACE2 is not highly expressed.

Hashimoto and colleagues then focused on the role of one particular neutral amino acid — tryptophan. A diet poor in tryptophan can induce pellagra, a vitamin-deficiency disease that affected Europe for centuries and became epidemic in the United States in the early twentieth century. Dietary improvement has now mostly restricted pellagra to some underdeveloped regions and to refugee camps. Pellagra is characterised by the 'four Ds': diarrhoea, dermatitis, dementia and, ultimately, death. Interestingly, patients with mutations in B^0AT1 can suffer from a pellagra-like syndrome[3].

To test for a direct connection between tryptophan transport and intestinal damage, Hashimoto *et al.* put normal mice on a low-tryptophan diet and found that they became highly susceptible to chemically induced intestinal inflammation, just as the ACE2-deficient mice did. Furthermore, the researchers found that supplementing the food of ACE2-deficient mice with a form of tryptophan that can be absorbed independently of B^0AT1 restored the animals' ability to resist chemically induced intestinal damage. Interestingly, the authors did not find enhanced susceptibility to chemical challenge in mice with a mutation that causes low levels of tryptophan owing to its loss through the urine, suggesting that resistance to inflammation is directly linked to local levels of tryptophan in the intestines.

To explain how ACE2 deficiency could exert effects in the large intestine despite the enzyme being expressed only at low levels in this gut region, Hashimoto *et al.* studied the role of antimicrobial peptides, which form a front-line defence mechanism against injury and infection. Antimicrobial peptides are secreted in large amounts by intestinal epithelial cells lining the walls of the intestine, and they travel throughout the gut, controlling the composition of intestinal bacterial communities[4] (Fig. 1). The researchers found that both ACE2-deficient mice and mice fed a low-tryptophan diet had reduced levels of antimicrobial peptides in their small intestines. Furthermore, when the authors analysed the microbiota in the large intestine of ACE2-deficient mice, they found that the species present diverged significantly from those in normal mice. Notably, these differences were reduced when the diet of the ACE2 mutant mice was supplemented with a source of tryptophan. In a definitive experiment, the investigators showed that when they transferred microbiota from ACE2-mutant mice or normal mice into germ-free mice, those that received microbiota from ACE2-mutant mice were more vulnerable to chemical damage. This result conclusively shows that the altered microbial community that developed in guts starved of tryptophan was the factor aggravating intestinal inflammation.

Although some aspects of this scenario had been proposed previously, Hashimoto *et al.* have integrated them into a coherent chain of events — from the genetic mutation through

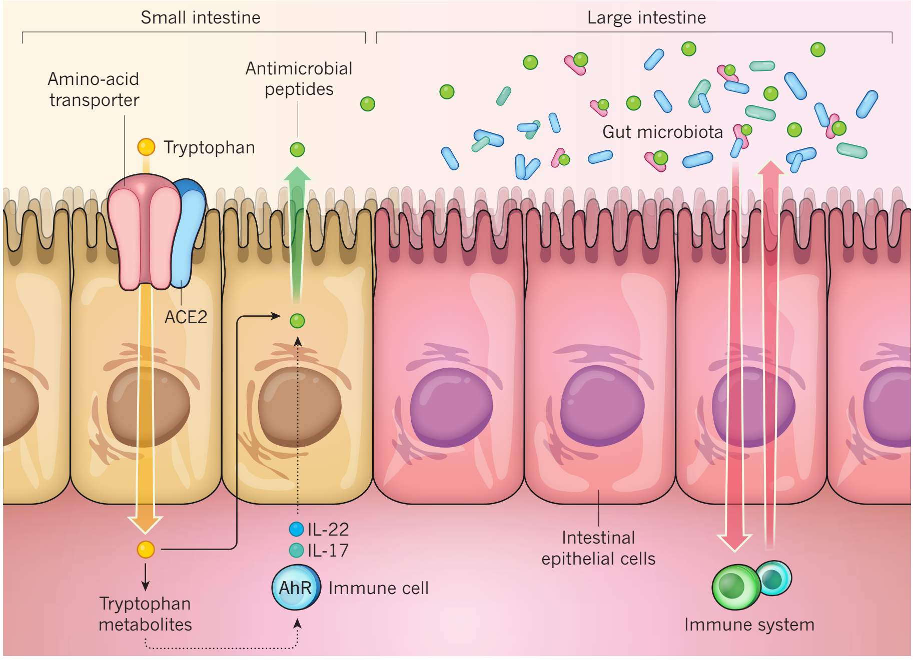


Figure 1 | A gut-friendly diet. The amino acid tryptophan is obtained from the diet and absorbed in the small intestine by a transporter protein that is dependent on the enzyme ACE2. Hashimoto *et al.*[2] demonstrate that mice that lack ACE2 or that are fed a diet lacking tryptophan have greater inflammation in the large intestine in response to damage than do mice with normal tryptophan levels. The tryptophan-deficient mice also have reduced levels of intestinal antimicrobial peptides, which are produced by intestinal epithelial cells and travel through the gut, controlling its microbial community. Gut microbes have a crucial role in immunity (red arrows), and the authors propose that altered microbiota in tryptophan-deficient mice is the cause of their susceptibility to inflammation. It is currently unknown whether the link between gut tryptophan levels and antimicrobial peptides established by Hashimoto and colleagues is direct (solid line) or involves other cell types (dotted line). For example, tryptophan and its metabolites can, through the action of the transcription-factor protein AhR (ref. 8), induce intestinal immune cells to produce the cytokine proteins IL-22 and IL-17, which then induce greater secretion of antimicrobial peptides.

specific nutrient deficiency to changes in the intestinal microbiota and increased susceptibility to inflammation. Their study also highlights the interdependency of key players in intestinal homeostasis, which include epithelial cells, the immune system and resident bacteria.

Several questions remain, however, particularly about the specific pathways that link tryptophan to inflammation. There is evidence that tryptophan and its derivatives have an important role in the immune response, although the mechanisms of this involvement are only partially unravelled. For example, low levels of tryptophan or other amino acids are directly sensed by a cellular pathway involving the protein mTOR (ref. 5), which plays a pivotal part in immunity through its effects on fundamental processes such as cell growth and transcription. Indeed, Hashimoto and colleagues found that mTOR activity was decreased in the intestines of their ACE2-deficient mice.

Tryptophan metabolism by enzymes such as tryptophan-2,3-dioxygenase (TDO)[6] and indoleamine-2,3-dioxygenase (IDO)[7], can also influence immune function. These enzymes generate tryptophan derivatives such as kynurenines, which can act as ligands for AhR (ref. 6), a transcription-factor protein that controls the expression of, among others, cytokine genes (cytokines are proteins involved in intercellular communication). In some immune cells, AhR activation induces the secretion of the cytokines IL-22 and IL-17 (ref. 8), which are sensed by intestinal epithelial cells, enhancing their production of antimicrobial peptides (Fig. 1). Indeed, diet-derived AhR ligands are essential for IL-22 production in the intestine, and deficiencies in this pathway lead to increased susceptibility to intestinal damage and inflammation[9,10]. These similarities between the effects of AhR deficiency and the characteristics of the ACE2-deficient mice in Hashimoto and colleagues' study make it tempting to speculate that AhR activation is the missing link between low dietary tryptophan and reduction in antimicrobial peptides, but the potential involvement of other tryptophan derivatives that have immune activity, such as serotonin[11], should not be neglected.

In summary, Hashimoto and colleagues have demonstrated that complex interactions involving diet, genetic background, host response and gut microbial ecology can be unravelled when using appropriate models. More studies dissecting these linked processes, and how they are regulated, will be crucial to

tackling public-health issues in economically disadvantaged areas where malnutrition-associated immunodeficiency and infectious diseases remain widespread. ■

Ana Izcue *is at the Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany and the Centre of Chronic Immunodeficiency, Freiburg.* **Fiona Powrie** *is in the Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 9DU, UK, and the Sir William Dunn School of Pathology, Oxford.*
e-mails: izcue@ie-freiburg.mpg.de; fiona.powrie@path.ox.ac.uk

1. Scrimshaw, N. S., Taylor, C. E. & Gordon, J. E. *Monogr. Ser. World Health Organ.* **57,** 3–329 (1968).
2. Hashimoto, T. *et al. Nature* **487,** 477–481 (2012).
3. Kleta, R. *et al. Nature Genet.* **36,** 999–1002 (2004).
4. Salzman, N. H. *et al. Nature Immunol.* **11,** 76–82 (2010).
5. Peter, C., Waldmann, H. & Cobbold, S. P. *Curr. Opin. Immunol.* **22,** 655–661 (2010).
6. Opitz, C. A. *et al. Nature* **478,** 197–203 (2011).
7. Munn, D. H. *et al. Science* **281,** 1191–1193 (1998).
8. Veldhoen, M. *et al. Nature* **453,** 106–109 (2008).
9. Li, Y. *et al. Cell* **147,** 629–640 (2011).
10. Kiss, E. A. *et al. Science* **334,** 1561–1565 (2011).
11. Ghia, J. E. *et al. Gastroenterology* **137,** 1649–1660 (2009).

HIV

Shock and kill

Antiretroviral therapies block HIV replication but they do not eliminate inactive viruses within cells. A clinical trial shows that a drug can revive HIV in patients as a potential first step towards a cure. SEE LETTER P.482

STEVEN G. DEEKS

HIV therapeutics is about to enter a new phase. Over the past 25 years, the focus has been almost entirely on developing and optimizing drugs aimed at inhibiting active HIV replication. Although this strategy has resulted in dramatic benefits for those with access to therapy, it has its limitations. Patients must take the drugs daily for life, and subtle toxicities accumulate over decades. Inflammation and immune dysfunction — which seem to have detrimental clinical consequences — persist even when viral replication is suppressed. Finally, and most importantly, the global resources necessary to deliver complex drug regimens, for many decades, to everyone in need are lacking. We require a therapeutic strategy in which a permanent disease-free state can be achieved after a more limited intervention. In other words, we need an effective cure for HIV infection[1,2]. On page 482 of this issue, Archin *et al.*[3] report a proof-of-concept study that provides the first evidence that such a cure might one day be feasible.

HIV persists during effective therapy in part because its genome becomes stably integrated in certain white blood cells known as resting memory CD4$^+$ T cells[4]. These latently infected cells do not express viral proteins and hence remain invisible to the immune system. If activated, however, they can ignite new rounds of viral replication — a risk that forces patients to remain on therapy indefinitely. Theoretically, drugs that reverse latency might lead sequentially to HIV RNA synthesis, viral protein production, release of HIV particles and (hopefully) killing of the infected cell by the virus or by the patient's immune system. Therefore, a cure might be possible if the latent virus in all infected cells can be forced out of its hiding place, leading ultimately to the death of the cells and to the elimination of the viral reservoir. Such a potential therapeutic approach is known as 'shock and kill' (Fig. 1).

Latency is maintained in part by the activity of histone deacetylase (HDAC), an enzyme that removes acetyl groups from DNA-bound histone proteins and, in so doing, affects gene expression. Although some HDAC inhibitors can induce mutations (at least *in vitro*), they might be able to reverse latency[5,6].

Archin *et al.* set out to test the anti-latency activity of vorinostat, an HDAC inhibitor approved for the management of certain cancers. Given safety concerns, the researchers first screened patients to ensure that they had an HIV reservoir that was responsive to the drug. To do this, the authors extracted white blood cells from the patients (by a procedure known as leukapheresis), purified resting memory CD4$^+$ T cells and then exposed such cells to the inhibitor. Of the 16 subjects screened, 11 exhibited a statistically significant vorinostat-mediated increase in HIV RNA expression and, of these, 8 patients eventually participated in the study.

The researchers administered a low dose of the drug (200 mg) to these eight subjects to ascertain tolerability. A few weeks later, a higher dose (400 mg) was given to determine anti-latency activity. Within six hours of this dose, the authors extracted resting memory CD4$^+$ T cells from the patients to measure the concentration of cell-associated HIV RNA. In the eight subjects, levels of HIV RNA in resting CD4$^+$ T cells increased in response to vorinostat, with the mean increase being 4.8-fold and the range 1.5–10.

As is common in such first-in-person clinical trials, this provocative study raises more questions than it answers. First, how should the field balance the ethical concerns about administering potentially toxic drugs to HIV-infected people who are otherwise healthy? The ideal population for these studies are those who have been doing well on long-term therapy, but this just happens to be the group with the lowest apparent need for a cure. Second, will future studies of anti-latency drugs require a costly and inconvenient leukapheresis before and after drug exposure? In San Francisco, the cost for such a procedure is over US$2,500.

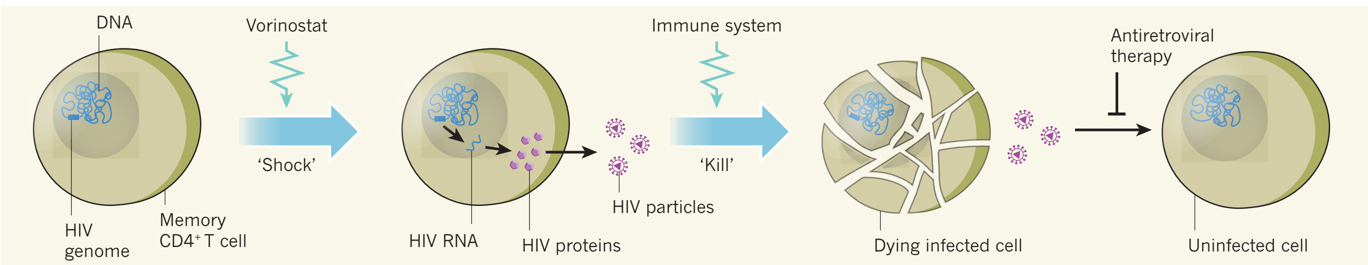


Figure 1 | Getting HIV out of its hiding place. Current treatments for HIV infection limit replication of the virus but do not eradicate it, as the viral genome remains integrated into the DNA of some white blood cells (memory CD4$^+$ T cells). Archin *et al.*[3] tested the potential for a new therapeutic approach ('shock and kill') that involves activating viral replication. The authors show that treating HIV-infected patients with the drug vorinostat leads to activation of HIV genes, as revealed by an increase in the synthesis of viral RNA. A similar approach could be one day used to 'awaken' the dormant HIV in patients; the infected cells would then be killed either by the virus itself (not shown) or by the patient's immune system. Moreover, current treatments such as highly active antiretroviral therapy would protect uninfected cells from becoming infected.

tackling public-health issues in economically disadvantaged areas where malnutrition-associated immunodeficiency and infectious diseases remain widespread. ■

Ana Izcue *is at the Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany and the Centre of Chronic Immunodeficiency, Freiburg.* **Fiona Powrie** *is in the Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 9DU, UK, and the Sir William Dunn School of Pathology, Oxford.*
e-mails: izcue@ie-freiburg.mpg.de; fiona.powrie@path.ox.ac.uk

1. Scrimshaw, N. S., Taylor, C. E. & Gordon, J. E. *Monogr. Ser. World Health Organ.* **57,** 3–329 (1968).
2. Hashimoto, T. *et al. Nature* **487,** 477–481 (2012).
3. Kleta, R. *et al. Nature Genet.* **36,** 999–1002 (2004).
4. Salzman, N. H. *et al. Nature Immunol.* **11,** 76–82 (2010).
5. Peter, C., Waldmann, H. & Cobbold, S. P. *Curr. Opin. Immunol.* **22,** 655–661 (2010).
6. Opitz, C. A. *et al. Nature* **478,** 197–203 (2011).
7. Munn, D. H. *et al. Science* **281,** 1191–1193 (1998).
8. Veldhoen, M. *et al. Nature* **453,** 106–109 (2008).
9. Li, Y. *et al. Cell* **147,** 629–640 (2011).
10. Kiss, E. A. *et al. Science* **334,** 1561–1565 (2011).
11. Ghia, J. E. *et al. Gastroenterology* **137,** 1649–1660 (2009).

HIV

Shock and kill

Antiretroviral therapies block HIV replication but they do not eliminate inactive viruses within cells. A clinical trial shows that a drug can revive HIV in patients as a potential first step towards a cure. SEE LETTER P.482

STEVEN G. DEEKS

HIV therapeutics is about to enter a new phase. Over the past 25 years, the focus has been almost entirely on developing and optimizing drugs aimed at inhibiting active HIV replication. Although this strategy has resulted in dramatic benefits for those with access to therapy, it has its limitations. Patients must take the drugs daily for life, and subtle toxicities accumulate over decades. Inflammation and immune dysfunction — which seem to have detrimental clinical consequences — persist even when viral replication is suppressed. Finally, and most importantly, the global resources necessary to deliver complex drug regimens, for many decades, to everyone in need are lacking. We require a therapeutic strategy in which a permanent disease-free state can be achieved after a more limited intervention. In other words, we need an effective cure for HIV infection[1,2]. On page 482 of this issue, Archin *et al.*[3] report a proof-of-concept study that provides the first evidence that such a cure might one day be feasible.

HIV persists during effective therapy in part because its genome becomes stably integrated in certain white blood cells known as resting memory $CD4^+$ T cells[4]. These latently infected cells do not express viral proteins and hence remain invisible to the immune system. If activated, however, they can ignite new rounds of viral replication — a risk that forces patients to remain on therapy indefinitely. Theoretically, drugs that reverse latency might lead sequentially to HIV RNA synthesis, viral protein production, release of HIV particles and (hopefully) killing of the infected cell by the virus or by the patient's immune system. Therefore, a cure might be possible if the latent virus in all infected cells can be forced out of its hiding place, leading ultimately to the death of the cells and to the elimination of the viral reservoir. Such a potential therapeutic approach is known as 'shock and kill' (Fig. 1).

Latency is maintained in part by the activity of histone deacetylase (HDAC), an enzyme that removes acetyl groups from DNA-bound histone proteins and, in so doing, affects gene expression. Although some HDAC inhibitors can induce mutations (at least *in vitro*), they might be able to reverse latency[5,6].

Archin *et al.* set out to test the anti-latency activity of vorinostat, an HDAC inhibitor approved for the management of certain cancers. Given safety concerns, the researchers first screened patients to ensure that they had an HIV reservoir that was responsive to the drug. To do this, the authors extracted white blood cells from the patients (by a procedure known as leukapheresis), purified resting memory $CD4^+$ T cells and then exposed such cells to the inhibitor. Of the 16 subjects screened, 11 exhibited a statistically significant vorinostat-mediated increase in HIV RNA expression and, of these, 8 patients eventually participated in the study.

The researchers administered a low dose of the drug (200 mg) to these eight subjects to ascertain tolerability. A few weeks later, a higher dose (400 mg) was given to determine anti-latency activity. Within six hours of this dose, the authors extracted resting memory $CD4^+$ T cells from the patients to measure the concentration of cell-associated HIV RNA. In the eight subjects, levels of HIV RNA in resting $CD4^+$ T cells increased in response to vorinostat, with the mean increase being 4.8-fold and the range 1.5–10.

As is common in such first-in-person clinical trials, this provocative study raises more questions than it answers. First, how should the field balance the ethical concerns about administering potentially toxic drugs to HIV-infected people who are otherwise healthy? The ideal population for these studies are those who have been doing well on long-term therapy, but this just happens to be the group with the lowest apparent need for a cure. Second, will future studies of anti-latency drugs require a costly and inconvenient leukapheresis before and after drug exposure? In San Francisco, the cost for such a procedure is over US$2,500.

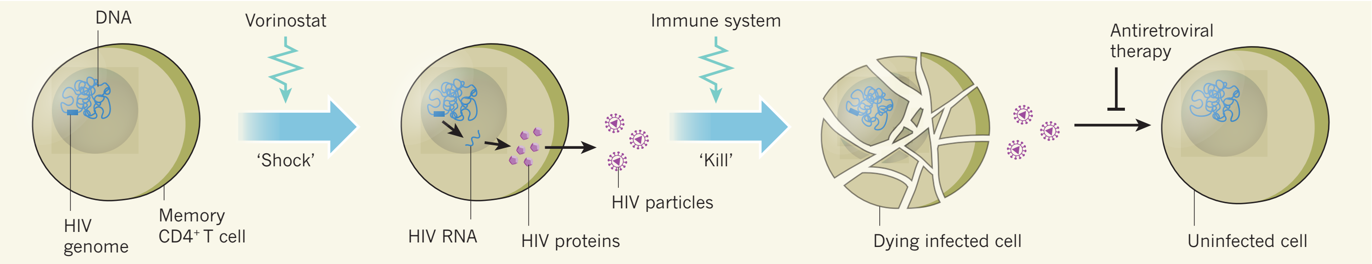


Figure 1 | Getting HIV out of its hiding place. Current treatments for HIV infection limit replication of the virus but do not eradicate it, as the viral genome remains integrated into the DNA of some white blood cells (memory $CD4^+$ T cells). Archin *et al.*[3] tested the potential for a new therapeutic approach ('shock and kill') that involves activating viral replication. The authors show that treating HIV-infected patients with the drug vorinostat leads to activation of HIV genes, as revealed by an increase in the synthesis of viral RNA. A similar approach could be one day used to 'awaken' the dormant HIV in patients; the infected cells would then be killed either by the virus itself (not shown) or by the patient's immune system. Moreover, current treatments such as highly active antiretroviral therapy would protect uninfected cells from becoming infected.

Therefore, a sensitive and high-throughput measure of viral latency is clearly needed.

Third, how much of the viral reservoir might be eliminated by HDAC inhibition? In the current study, the cells isolated from 5 of the 16 tested subjects failed to show any response to vorinostat and therefore these patients were not eligible for the clinical trial. Although those who eventually received the drug exhibited an increase in HIV RNA production, the overall clinical effect was probably marginal, as the concentration of viral particles in plasma did not increase and there was no apparent decrease in the size of the HIV reservoir.

Fourth, which assays will we use in the future to screen potential drug candidates for anti-latency activity? Vorinostat has demonstrated activity in most tests[5,6], but not all[7]. Finally, what is the fate of the virus-producing cells after HDAC inhibition? Although many investigators have assumed that either the virus or the host immune system would destroy such cells and would therefore clear the virus, recent data suggest that this might not be true[8]. If the ultimate goal is to shock and kill infected cells, then more work is needed on how to eliminate them[9,10].

Despite these uncertainties, the importance of Archin and colleagues' study cannot be overstated, as it provides a rationale for an entirely new approach to the management of HIV infection. In many ways, their results are comparable to the initial finding that the anti-HIV drug zidovudine (also known as AZT) reduced viral replication in people. Even though zidovudine proved to have limited benefit on its own, the result showed that HIV could be inhibited, and the drug eventually became the basis for the first generation of drug combination regimens. I hope that HDAC inhibitors could ultimately become part of a combination approach to curing HIV infection. ■

Steven G. Deeks *is in the Department of Medicine, University of California, San Francisco, San Francisco, California 94110, USA.*
e-mail: sdeeks@php.ucsf.edu

1. Richman, D. D. *et al. Science* **323,** 1304–1307 (2009).
2. Trono, D. *et al. Science* **329,** 174–180 (2010).
3. Archin, N. M. *et al. Nature* **487,** 482–485 (2012).
4. Siliciano, J. D. *et al. Nature Med.* **9,** 727–728 (2003).
5. Archin, N. M. *et al. AIDS Res. Hum. Retroviruses* **25,** 207–212 (2009).
6. Contreras, X. *et al. J. Biol. Chem.* **284,** 6782–6789 (2009).
7. Blazkova, J. *et al. J. Infect. Dis.* http://dx.doi.org/10.1093/infdis/jis412 (2012).
8. Shan, L. *et al. Immunity* **36,** 491–501 (2012).
9. Hansen, S. G. *et al. Nature* **473,** 523–527 (2011).
10. Persaud, D. *et al. AIDS* **25,** 2227–2234 (2011).

The author declares competing financial interests. See go.nature.com/fbngjd for details.

MATERIALS SCIENCE

The abilities of instabilities

A method has been developed for preparing a variety of potentially useful spherical particles, ranging from several nanometres to millimetres in diameter. It relies on the same fluid instability that causes taps to drip. **See Letter p.463**

ALI PASSIAN & THOMAS THUNDAT

Despite the many advantages offered by nanoparticles as materials for a wide variety of applications, their use is currently limited by the lack of scalable, high-throughput techniques for synthesizing multifunctional particles of uniform size. Nanoparticles that are made from multiple materials are especially interesting, because they have potential applications in many devices, such as biological and chemical sensors. On page 463 of this issue, Kaufman *et al.*[1] report a scalable fabrication technique that uses the dynamic instabilities of a fluid column to make particles of complex architecture*. Using this approach, they have prepared same-sized particles from combinations of vastly different materials, and with an unprecedented range of particle sizes.

Instabilities in fluid systems often produce undesirable effects — for example, standing up in a canoe can induce buoyancy instabilities that capsize the boat. However, instabilities can prove highly useful if they are harnessed properly, as Kaufman *et al.* demonstrate. Their fabrication technique depends on 'Plateau–Rayleigh instabilities' (PRIs), which are often manifested as remarkable flow patterns when a fluid is perturbed by heat.

Although the effects of PRIs can be quite subtle, these instabilities can grow exponentially, spatially and temporally modifying the motion or the flow pattern of a continuous body of fluid until it breaks up into 'pieces'. This is what happens when droplets form from dripping taps. The resulting fluid domains tend to form as spheres, as demonstrated in the microgravity environments of space laboratories[2]. When small sinusoidal perturbations are engendered along a fluid column composed of different liquids, the wavelengths and growth rates of the perturbations depend on the fluid viscosities and on the diameters of the participating fluid columns — an effect exploited by Kaufman *et al.* in their technique.

The authors' process for making particles begins with a 'preform' — a centimetre-scale rod in which a core of the materials to be made into particles is clad by a shell of scaffold material. This rod is heated and drawn into a fibre, until the diameter of the core matches the desired diameter of the particles. The fibre is then heated in such a way as to induce PRIs, which cause the core to split into spherical droplets (Fig. 1). These droplets are frozen *in situ*, forming particles encased within the scaffold material.

Microstructure-fabrication techniques in which PRIs break up multilayer, coaxial fluidic columns have been explored previously. For example, microspheres[3] have been prepared using a microfluidic device in which droplets of a gel flowing at the core of a column of oil are polymerized by exposure to ultraviolet light. A similar approach[4] has been used to fabricate microcapsules and strips of connected microcapsules. However, Kaufman and colleagues' method differs by virtue of having achieved PRIs in prefabricated core–shell fibres. Most impressively, their technique is scalable, allows a variety of particle morphologies to be made (including 'beach ball' particles that contain six segments made of alternating materials), and works for particle sizes ranging from 20 nanometres to 2 millimetres — a whopping five orders of magnitude.

Other methods for making particles — based either on chemical synthesis[5] or on physical methods such as laser ablation in liquids[6] — have been used to generate suspensions of a variety of highly pure nanoparticles. However, a great advantage of Kaufman and colleagues' technique is its ability to create multilayered nanoparticles.

Understanding and controlling fluid behaviour at a small scale is of tremendous technological importance. The Navier–Stokes equation is used to describe the motion of fluids, but this must often be modified to take into account the constraints of specific flow systems. For example, a study[7] of the de-wetting of thin polymer films has shown that thermal noise — random forces in a fluid caused by molecular motion that is induced by thermal agitation — can strongly influence the time- and length-scales of the films' flow. Given that Kaufman *et al.* use thermal perturbation as a means to induce instabilities, and that thermal noise increases with temperature, one would expect the break-up of the thin liquid columns in their technique to be influenced by such

*This article and the paper under discussion[1] were published online on 18 July 2012.

Therefore, a sensitive and high-throughput measure of viral latency is clearly needed.

Third, how much of the viral reservoir might be eliminated by HDAC inhibition? In the current study, the cells isolated from 5 of the 16 tested subjects failed to show any response to vorinostat and therefore these patients were not eligible for the clinical trial. Although those who eventually received the drug exhibited an increase in HIV RNA production, the overall clinical effect was probably marginal, as the concentration of viral particles in plasma did not increase and there was no apparent decrease in the size of the HIV reservoir.

Fourth, which assays will we use in the future to screen potential drug candidates for anti-latency activity? Vorinostat has demonstrated activity in most tests[5,6], but not all[7]. Finally, what is the fate of the virus-producing cells after HDAC inhibition? Although many investigators have assumed that either the virus or the host immune system would destroy such cells and would therefore clear the virus, recent data suggest that this might not be true[8]. If the ultimate goal is to shock and kill infected cells, then more work is needed on how to eliminate them[9,10].

Despite these uncertainties, the importance of Archin and colleagues' study cannot be overstated, as it provides a rationale for an entirely new approach to the management of HIV infection. In many ways, their results are comparable to the initial finding that the anti-HIV drug zidovudine (also known as AZT) reduced viral replication in people. Even though zidovudine proved to have limited benefit on its own, the result showed that HIV could be inhibited, and the drug eventually became the basis for the first generation of drug combination regimens. I hope that HDAC inhibitors could ultimately become part of a combination approach to curing HIV infection. ■

Steven G. Deeks *is in the Department of Medicine, University of California, San Francisco, San Francisco, California 94110, USA.*
e-mail: sdeeks@php.ucsf.edu

1. Richman, D. D. *et al. Science* **323,** 1304–1307 (2009).
2. Trono, D. *et al. Science* **329,** 174–180 (2010).
3. Archin, N. M. *et al. Nature* **487,** 482–485 (2012).
4. Siliciano, J. D. *et al. Nature Med.* **9,** 727–728 (2003).
5. Archin, N. M. *et al. AIDS Res. Hum. Retroviruses* **25,** 207–212 (2009).
6. Contreras, X. *et al. J. Biol. Chem.* **284,** 6782–6789 (2009).
7. Blazkova, J. *et al. J. Infect. Dis.* http://dx.doi.org/10.1093/infdis/jis412 (2012).
8. Shan, L. *et al. Immunity* **36,** 491–501 (2012).
9. Hansen, S. G. *et al. Nature* **473,** 523–527 (2011).
10. Persaud, D. *et al. AIDS* **25,** 2227–2234 (2011).

The author declares competing financial interests.
See go.nature.com/fbngjd for details.

MATERIALS SCIENCE

The abilities of instabilities

A method has been developed for preparing a variety of potentially useful spherical particles, ranging from several nanometres to millimetres in diameter. It relies on the same fluid instability that causes taps to drip. **SEE LETTER P.463**

ALI PASSIAN & THOMAS THUNDAT

Despite the many advantages offered by nanoparticles as materials for a wide variety of applications, their use is currently limited by the lack of scalable, high-throughput techniques for synthesizing multifunctional particles of uniform size. Nanoparticles that are made from multiple materials are especially interesting, because they have potential applications in many devices, such as biological and chemical sensors. On page 463 of this issue, Kaufman *et al.*[1] report a scalable fabrication technique that uses the dynamic instabilities of a fluid column to make particles of complex architecture*. Using this approach, they have prepared same-sized particles from combinations of vastly different materials, and with an unprecedented range of particle sizes.

Instabilities in fluid systems often produce undesirable effects — for example, standing up in a canoe can induce buoyancy instabilities that capsize the boat. However, instabilities can prove highly useful if they are harnessed properly, as Kaufman *et al.* demonstrate. Their fabrication technique depends on 'Plateau–Rayleigh instabilities' (PRIs), which are often manifested as remarkable flow patterns when a fluid is perturbed by heat.

Although the effects of PRIs can be quite subtle, these instabilities can grow exponentially, spatially and temporally modifying the motion or the flow pattern of a continuous body of fluid until it breaks up into 'pieces'. This is what happens when droplets form from dripping taps. The resulting fluid domains tend to form as spheres, as demonstrated in the microgravity environments of space laboratories[2]. When small sinusoidal perturbations are engendered along a fluid column composed of different liquids, the wavelengths and growth rates of the perturbations depend on the fluid viscosities and on the diameters of the participating fluid columns — an effect exploited by Kaufman *et al.* in their technique.

The authors' process for making particles begins with a 'preform' — a centimetre-scale rod in which a core of the materials to be made into particles is clad by a shell of scaffold material. This rod is heated and drawn into a fibre, until the diameter of the core matches the desired diameter of the particles. The fibre is then heated in such a way as to induce PRIs, which cause the core to split into spherical droplets (Fig. 1). These droplets are frozen *in situ*, forming particles encased within the scaffold material.

Microstructure-fabrication techniques in which PRIs break up multilayer, coaxial fluidic columns have been explored previously. For example, microspheres[3] have been prepared using a microfluidic device in which droplets of a gel flowing at the core of a column of oil are polymerized by exposure to ultraviolet light. A similar approach[4] has been used to fabricate microcapsules and strips of connected microcapsules. However, Kaufman and colleagues' method differs by virtue of having achieved PRIs in prefabricated core–shell fibres. Most impressively, their technique is scalable, allows a variety of particle morphologies to be made (including 'beach ball' particles that contain six segments made of alternating materials), and works for particle sizes ranging from 20 nanometres to 2 millimetres — a whopping five orders of magnitude.

Other methods for making particles — based either on chemical synthesis[5] or on physical methods such as laser ablation in liquids[6] — have been used to generate suspensions of a variety of highly pure nanoparticles. However, a great advantage of Kaufman and colleagues' technique is its ability to create multilayered nanoparticles.

Understanding and controlling fluid behaviour at a small scale is of tremendous technological importance. The Navier–Stokes equation is used to describe the motion of fluids, but this must often be modified to take into account the constraints of specific flow systems. For example, a study[7] of the de-wetting of thin polymer films has shown that thermal noise — random forces in a fluid caused by molecular motion that is induced by thermal agitation — can strongly influence the time- and length-scales of the films' flow. Given that Kaufman *et al.* use thermal perturbation as a means to induce instabilities, and that thermal noise increases with temperature, one would expect the break-up of the thin liquid columns in their technique to be influenced by such

*This article and the paper under discussion[1] were published online on 18 July 2012.

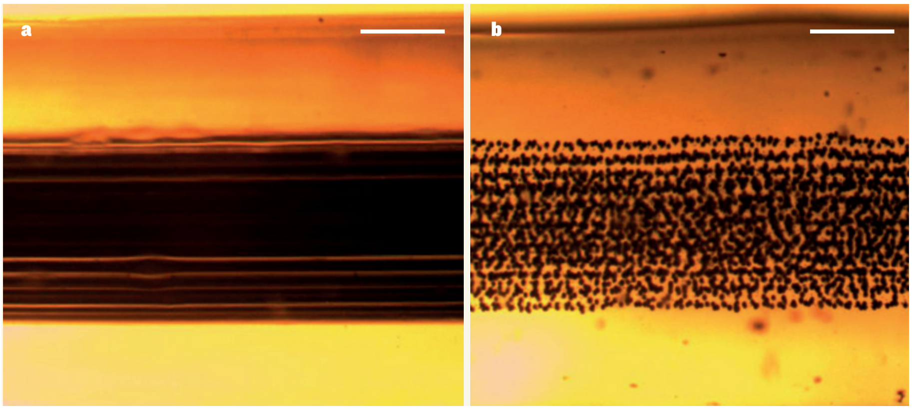


Figure 1 | An orderly break-up. Kaufman *et al.*[1] report a technique for making particles from fibres of material clad by a polymeric scaffold. **a**, In this image, black fibres of a glass (arsenic triselenide) are visible at the centre of an amber-coloured column of polymer (polyethersulphone). **b**, When heated in such a way as to generate fluid instabilities in the glass, the fibres break up into equally sized, spherical particles. Scale bar, 200 micrometres.

noise. The authors' theoretical description[1] of fluid flow in their system might therefore benefit from the inclusion of prevailing stochastic processes, as this would enable them to obtain better design parameters.

Furthermore, given that the surface tension of fluids generally decreases with increasing temperature, and that fluids move from regions of low surface tension to regions of high surface tension, it is worth considering whether the flow in Kaufman and colleagues' system receives additional contributions from temperature gradients generated in their experiments. Instabilities and flow systems driven by surface-tension gradients find applications in processes that involve the self-assembly of nanoparticles[8], and in the optical translation of droplets[9]. They also give rise to fascinating phenomena such as 'tears of wine' — the clear ring of liquid that forms on the inside of a filled wine glass above the drink, from which droplets constantly form and trickle back into the wine. In the case of PRIs, when a droplet breaks away from a column of fluid, the newly formed meniscus undergoes oscillations. This effect is important, for example in the operation of inkjet printers[10].

A potential limitation of Kaufman and colleagues' method is the nature of the materials that can be used in the nanoparticle fabrication. Chemically reactive and/or thermally incompatible materials (those that become incompatible when heated, for example because one of the materials degrades or changes its physical state) might not be able to sustain sufficient columnar or interfacial integrity for the PRIs to cause the desired break-up. This could also make it difficult to remove particles from the surrounding polymer, unless they can be dissolved in another chemical — although this runs the risk of the nanoparticle population being contaminated by chemical residues. Such contamination can affect the surface reactivity of nanoparticles, which, in turn, can adversely affect their functions in applications, for example when used in sensors.

Although not applicable to every material, Kaufman and colleagues' technique is ideal for making composite spherical particles from selected materials, such as the beach ball particles mentioned earlier. The authors also synthesized particles consisting of a core of one material covered with a shell of another, and 'Janus' particles, which have hemispheres made of different materials. Indeed, the authors' technique overcomes the limitations of present methods for making Janus particles, which modify only the particles' surfaces. Such particles are especially interesting because they can be used in sensors, actuators and energy-conversion devices, and as building blocks for self-assembling structures. The ability to make them reliably therefore opens up new vistas in the development of these applications. ■

Ali Passian *is in the Measurement Science and Systems Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.* **Thomas Thundat** *is in the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada.*
e-mails: passianan@ornl.gov; thundat@ualberta.ca

1. Kaufman, J. J. *et al. Nature* **487,** 463–467 (2012).
2. www.physicscentral.com/explore/sots/episode5.cfm
3. Jeong, W. J. *et al. Langmuir* **21,** 3738–3741 (2005).
4. Oh, H. J., Kim, S. H., Baek, J. Y., Seong, G. H. & Lee, S. H. *J. Micromech. Microeng.* **16,** 285–291 (2006).
5. Schmidt, H. *Appl. Organometal. Chem.* **15,** 331–343 (2001).
6. Barcikowski, S. & Mafuné, F. *J. Phys. Chem. C* **115,** 4985–5180 (2011).
7. Fetzer, R., Rauscher, M., Seemann, R., Jacobs, K. & Mecke, K. *Phys. Rev. Lett.* **99,** 114503 (2007).
8. Cai, Y. & Zhang Newby, B. *J. Am. Chem. Soc.* **130,** 6076–6077 (2008).
9. Passian, A. *et al. Phys. Rev. E* **73,** 066311 (2006).
10. Farahi, R. H. *Opt. Lett.* **34,** 3148–3150 (2009).

VISION

Looking to develop sight

Producing a single image from two eyes requires complex brain circuitry. A comparison of neural responses in babies shows that early visual stimulation following premature birth leads to accelerated development of the visual system.

EILEEN BIRCH

Our eyes look at the world from two different positions, yet we see one unified three-dimensional environment. The brain extracts information from the two-dimensional images formed on the two retinas by matching corresponding points, and then reconstructs the third dimension using residual disparities between the monocular images. When and how the brain develops the capacity for this complex binocular task has long been a focus of research for neuroscientists and vision researchers. Infants of both humans and non-human primates exhibit a sudden post-natal onset of binocular vision that corresponds to the start of the 'critical period', during which binocular function can be severely and permanently disrupted by abnormal visual experience. This maturational time course led to a consensus view that visual development during the pre-critical period is guided by genetically programmed molecular and neuronal signals, whereas environmental signals maintain and refine the neuronal circuitry for binocular vision during the critical period. Writing in *Proceedings of the National Academy of Sciences*, Jandó *et al.*[1] challenge this consensus and offer the first evidence in human infants that visual experience also guides development during the pre-critical period.

Interest in using visual-system development as a model for studying the role of experience in guiding brain development was stimulated by Hubel and Wiesel's elegant descriptions[2,3] in the early 1960s of the organization of neural circuits in the primary visual cortex — the first visual area of the cortex where information from the retinas is processed. A seminal contribution from these researchers was the identification of a feature of neural organization that is crucial for combining information

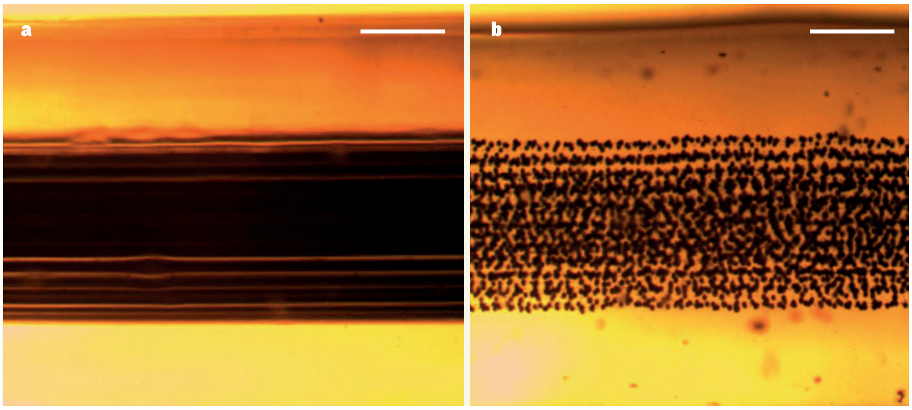


Figure 1 | An orderly break-up. Kaufman *et al.*[1] report a technique for making particles from fibres of material clad by a polymeric scaffold. **a**, In this image, black fibres of a glass (arsenic triselenide) are visible at the centre of an amber-coloured column of polymer (polyethersulphone). **b**, When heated in such a way as to generate fluid instabilities in the glass, the fibres break up into equally sized, spherical particles. Scale bar, 200 micrometres.

noise. The authors' theoretical description[1] of fluid flow in their system might therefore benefit from the inclusion of prevailing stochastic processes, as this would enable them to obtain better design parameters.

Furthermore, given that the surface tension of fluids generally decreases with increasing temperature, and that fluids move from regions of low surface tension to regions of high surface tension, it is worth considering whether the flow in Kaufman and colleagues' system receives additional contributions from temperature gradients generated in their experiments. Instabilities and flow systems driven by surface-tension gradients find applications in processes that involve the self-assembly of nanoparticles[8], and in the optical translation of droplets[9]. They also give rise to fascinating phenomena such as 'tears of wine' — the clear ring of liquid that forms on the inside of a filled wine glass above the drink, from which droplets constantly form and trickle back into the wine. In the case of PRIs, when a droplet breaks away from a column of fluid, the newly formed meniscus undergoes oscillations. This effect is important, for example in the operation of inkjet printers[10].

A potential limitation of Kaufman and colleagues' method is the nature of the materials that can be used in the nanoparticle fabrication. Chemically reactive and/or thermally incompatible materials (those that become incompatible when heated, for example because one of the materials degrades or changes its physical state) might not be able to sustain sufficient columnar or interfacial integrity for the PRIs to cause the desired break-up. This could also make it difficult to remove particles from the surrounding polymer, unless they can be dissolved in another chemical — although this runs the risk of the nanoparticle population being contaminated by chemical residues. Such contamination can affect the surface reactivity of nanoparticles, which, in turn, can adversely affect their functions in applications, for example when used in sensors.

Although not applicable to every material, Kaufman and colleagues' technique is ideal for making composite spherical particles from selected materials, such as the beach ball particles mentioned earlier. The authors also synthesized particles consisting of a core of one material covered with a shell of another, and 'Janus' particles, which have hemispheres made of different materials. Indeed, the authors' technique overcomes the limitations of present methods for making Janus particles, which modify only the particles' surfaces. Such particles are especially interesting because they can be used in sensors, actuators and energy-conversion devices, and as building blocks for self-assembling structures. The ability to make them reliably therefore opens up new vistas in the development of these applications. ■

Ali Passian *is in the Measurement Science and Systems Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.* **Thomas Thundat** *is in the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada.*
e-mails: passianan@ornl.gov; thundat@ualberta.ca

1. Kaufman, J. J. *et al. Nature* **487,** 463–467 (2012).
2. www.physicscentral.com/explore/sots/episode5.cfm
3. Jeong, W. J. *et al. Langmuir* **21,** 3738–3741 (2005).
4. Oh, H. J., Kim, S. H., Baek, J. Y., Seong, G. H. & Lee, S. H. *J. Micromech. Microeng.* **16,** 285–291 (2006).
5. Schmidt, H. *Appl. Organometal. Chem.* **15,** 331–343 (2001).
6. Barcikowski, S. & Mafuné, F. *J. Phys. Chem. C* **115,** 4985–5180 (2011).
7. Fetzer, R., Rauscher, M., Seemann, R., Jacobs, K. & Mecke, K. *Phys. Rev. Lett.* **99,** 114503 (2007).
8. Cai, Y. & Zhang Newby, B. *J. Am. Chem. Soc.* **130,** 6076–6077 (2008).
9. Passian, A. *et al. Phys. Rev. E* **73,** 066311 (2006).
10. Farahi, R. H. *Opt. Lett.* **34,** 3148–3150 (2009).

VISION

Looking to develop sight

Producing a single image from two eyes requires complex brain circuitry. A comparison of neural responses in babies shows that early visual stimulation following premature birth leads to accelerated development of the visual system.

EILEEN BIRCH

Our eyes look at the world from two different positions, yet we see one unified three-dimensional environment. The brain extracts information from the two-dimensional images formed on the two retinas by matching corresponding points, and then reconstructs the third dimension using residual disparities between the monocular images. When and how the brain develops the capacity for this complex binocular task has long been a focus of research for neuroscientists and vision researchers. Infants of both humans and non-human primates exhibit a sudden post-natal onset of binocular vision that corresponds to the start of the 'critical period', during which binocular function can be severely and permanently disrupted by abnormal visual experience. This maturational time course led to a consensus view that visual development during the pre-critical period is guided by genetically programmed molecular and neuronal signals, whereas environmental signals maintain and refine the neuronal circuitry for binocular vision during the critical period. Writing in *Proceedings of the National Academy of Sciences*, Jandó *et al.*[1] challenge this consensus and offer the first evidence in human infants that visual experience also guides development during the pre-critical period.

Interest in using visual-system development as a model for studying the role of experience in guiding brain development was stimulated by Hubel and Wiesel's elegant descriptions[2,3] in the early 1960s of the organization of neural circuits in the primary visual cortex — the first visual area of the cortex where information from the retinas is processed. A seminal contribution from these researchers was the identification of a feature of neural organization that is crucial for combining information

from the two eyes to accurately reconstruct the three-dimensional world. They described how nerve axons from the thalamus — the brain's primary relay centre for visual information received from the retina — fan out into a broad band and terminate in the visual cortex in alternating eye-specific zones called cortical ocular dominance columns.

Hubel and Wiesel also performed a series of simple but elegant experiments in cats and primates in which they showed that stitching one eye closed during early visual development permanently altered the structure of the columns, shifting the ocular dominance of cortical cells to the non-deprived eye[2,3]. They hypothesized that visual experience guides the development of ocular dominance columns during a critical period shortly, but not immediately, after birth. As a result, the developing neuronal connectivity and function is susceptible to disruption by visual deprivation or abnormal visual experience during this period.

For the next three and a half decades, neuroscientists regarded this hypothesis as a fundamental principle of the mechanisms guiding the emergence of the complex and precise organization of functionally specialized neuronal circuits — in vision and in other neural pathways. In the past decade, however, this dogma has been challenged. We now know that ocular dominance columns form much earlier after birth, during the pre-critical period[4,5]. In fact, there is evidence that at least a protomap of ocular dominance columns develops prenatally[5]. Nonetheless, early formation of ocular dominance columns does not mean that their development is pre-programmed. There is evidence that, during the pre-critical period, genetically encoded cell-patterning mechanisms and visual experience can influence the formation of balanced ocular-dominance maps[6,7].

Investigations into the role of visual experience in refining ocular dominance columns during the pre-critical period have largely been limited to experiments in which the visual cortex is deprived of extrinsic stimulation during development. However, such experiments are ambiguous, because there is a high level of intrinsic neuronal activity in the visual cortex that can be altered by visual deprivation. Similarly, elimination of intrinsic stimulation might alter any interplay between intrinsic and extrinsic factors that contribute to the development of neuronal circuits. Jandó *et al.*[1] are the first to examine the effect of additional visual experience, rather than deprivation, on the development of binocular function.

To do this, the authors compared the age of onset of a binocular response from the visual cortex in human infants born either prematurely or at term. They reasoned that pre-programmed development of binocular cortical function would lead to the age of onset in preterm infants being the same as that in term infants when the preterm infants' ages were adjusted to reflect the age the child would be if

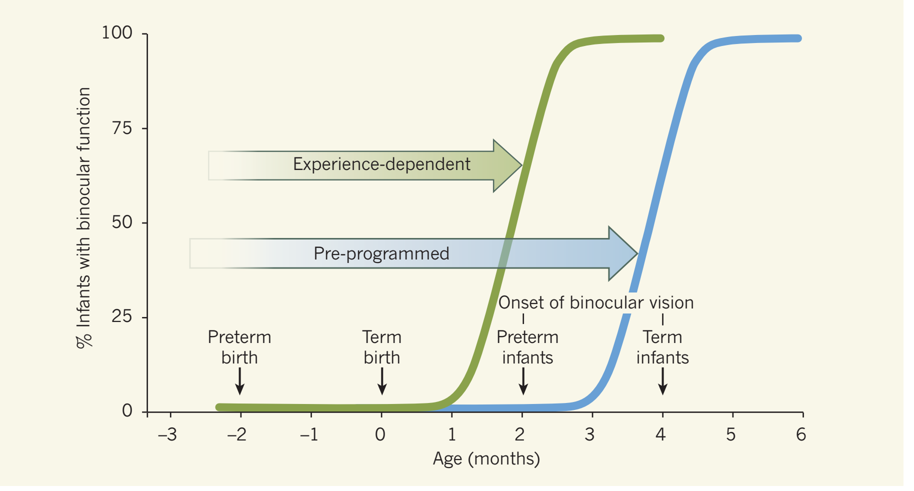


Figure 1 | Experience counts. Binocular vision is absent at birth and seems to switch on when a baby is about four months old. Jandó *et al.*[1] monitored the activity of the visual cortex region of the brains of babies who were born at term (nine months post-conception) and babies who were born two months prematurely, to determine when they first displayed a binocularly evoked response. Binocular vision developed at four months after birth in both cohorts. This suggests that four months of post-natal visual experience is the key determinant of the onset of binocular vision, even in premature babies (green curve). This is in contrast to the alternative scenario of a pre-programmed onset age that is guided by genetically encoded molecular and neuronal signals. In this case, additional visual experience should not alter the rate of maturation, and binocular onset would be expected to occur six months post-natally in the premature babies (blue curve).

the pregnancy had gone to term. Alternatively, a role for experience would be indicated by an equivalent age of onset when age is expressed in post-natal months — the number of months of visual experience (Fig. 1).

To distinguish between these possibilities, Jandó and colleagues presented random dot patterns — either the same pattern to both eyes or a different pattern to each — to babies who were born at term (nine months post-conception) or two months early (seven months post-conception). The researchers monitored the babies' brain activity about once a month post-natally to determine when the infants first responded differently to patterns that were matched versus patterns that did not match. The answer was clear-cut — the onset of the binocular response occurred 4 months after birth for both term and preterm infants.

This finding indicates a strong role for visual experience in guiding the development of binocular organization of the visual cortex during the pre-critical period. The current literature supports a close association between the onset of binocular function and the onset of the critical period, but it remains to be determined whether the precocious onset of binocular function in preterm infants also heralds the transition from pre-critical period to critical period.

If the transition from the pre-critical to the critical period is advanced by preterm birth, we may need to rethink current approaches to the treatment of disorders that can disrupt the development of binocular vision during infancy, such as esotropia (crossed eyes) and anisometropia (different refractive errors in each eye). These two disorders are far more prevalent among preterm infants than among infants born at term[8]. The current therapeutic focus is on early surgery or optical treatment to minimize the duration of abnormal visual experience during the critical period. However, Jandó and colleagues' finding that the early visual experience afforded by preterm birth can accelerate the onset of binocular vision raises the possibility that preterm infants could require intervention timed to their post-natal age — not their adjusted age — to prevent permanent changes in the structural and functional organization of the visual cortex. ■

Eileen Birch *is in the Pediatric Laboratory, Retina Foundation of the Southwest, Dallas, 75231 Texas, USA.*
e-mail: ebirch@retinafoundation.org

1. Jandó, G. *et al. Proc. Natl Acad. Sci. USA* http://dx.doi.org/10.1073/pnas.1203096109 (2012).
2. Hubel, D. H. *Nature* **299,** 515–524 (1982).
3. Wiesel, T. N. *Nature* **299,** 583–591 (1982).
4. Horton, J. C. & Hocking, D. R. *J. Neurosci.* **16,** 1791–1807 (1996).
5. Rakic, P. *Nature* **261,** 467–471 (1976).
6. Levelt, C. N. & Hübener, M. *Ann. Rev. Neurosci.* **35,** 309–330 (2012).
7. Feller, M. B. & Scanziani, M. *Curr. Opin. Neurobiol.* **15,** 94–100 (2005).
8. O'Connor, A. R. *et al. Pediatrics* **109,** 12–18 (2002).

CORRECTION

In the News & Views article 'The ancestral dinner table' by Margaret Schoeninger (*Nature* **487,** 42–43; 2012), reference 7 should have been cited in the fifth paragraph, rather than reference 4, in the sentence "This predominant C_4 signal led some researchers to suggest that...".

doi:10.1038/nature11301

Alignment of the stellar spin with the orbits of a three-planet system

Roberto Sanchis-Ojeda[1], Daniel C. Fabrycky[2], Joshua N. Winn[1], Thomas Barclay[3,4], Bruce D. Clarke[4,5], Eric B. Ford[6], Jonathan J. Fortney[2], John C. Geary[7], Matthew J. Holman[7], Andrew W. Howard[8], Jon M. Jenkins[4,5], David Koch[4], Jack J. Lissauer[4], Geoffrey W. Marcy[8], Fergal Mullally[4,5], Darin Ragozzine[7], Shawn E. Seader[4,5], Martin Still[3,4] & Susan E. Thompson[4,5]

The Sun's equator and the planets' orbital planes are nearly aligned, which is presumably a consequence of their formation from a single spinning gaseous disk. For exoplanetary systems this well-aligned configuration is not guaranteed: dynamical interactions may tilt planetary orbits, or stars may be misaligned with the protoplanetary disk through chaotic accretion[1] , magnetic interactions[2] or torques from neighbouring stars. Indeed, isolated 'hot Jupiters' are often misaligned and even orbiting retrograde[3,4]. Here we report an analysis of transits of planets over starspots[5–7] on the Sun-like star Kepler-30 (ref. 8), and show that the orbits of its three planets are aligned with the stellar equator. Furthermore, the orbits are aligned with one another to within a few degrees. This configuration is similar to that of our Solar System, and contrasts with the isolated hot Jupiters. The orderly alignment seen in the Kepler-30 system suggests that high obliquities are confined to systems that experienced disruptive dynamical interactions. Should this be corroborated by observations of other coplanar multi-planet systems, then star–disk misalignments would be ruled out as the explanation for the high obliquities of hot Jupiters, and dynamical interactions would be implicated as the origin of hot Jupiters.

Kepler-30 is a star of nearly solar mass and radius, but it is probably younger than the Sun, judging from its faster rotation and more prominent starspots[8]. The starspots are crucial to measuring the stellar obliquity (the angle between the rotational and orbital angular momentum vectors). Starspots produce two effects: quasi-periodic variation (QPV) in flux caused by rotation, and shorter-term 'anomalies' in flux caused by the transit of a planet in front of a spot. The obliquity can be measured if one observes a sequence of anomalies[5,7], or a few single anomalies and the accompanying QPV[6], as long as the effects of a single spot or compact group of spots can be isolated. This technique has been previously applied to solitary short-period planets, but not longer-period planets or systems of multiple planets. The other widely used technique for measuring stellar obliquities, the Rossiter–McLaughlin effect[9], relies on precise spectroscopy during transits and would be impractical for a star as faint as Kepler-30.

We analysed 2.5 years of nearly continuous photometric time-series data from the Kepler space telescope[10]. The data set includes 27 transits of Kepler-30b ('planet b'; orbital period, ~29 days; radius, ~4 R_E, where R_E is the Earth's radius), 12 transits of Kepler-30c ('planet c'; 60 days; 13R_E), and 5 transits of Kepler-30d ('planet d'; 143 days; 10R_E). After removing instrumental artefacts (see Supplementary Information), we detected QPV with an amplitude (peak-to-peak) of 1.5%. The stellar rotation period is 16.0 ± 0.4 days, based on a Lomb–Scargle periodogram[11] (Supplementary Information).

To enable the obliquity analysis, we searched for anomalies during transits that are large enough in amplitude and long enough in duration to be caused by the same starspots that produce the QPV. Many such anomalies were identified during transits of the largest planet, c. A strong correlation exists between the timing of the anomaly relative to mid-transit, and the phase of the QPV: anomalies observed near mid-transit are found when the QPV is near a local minimum, whereas anomalies occurring before (or after) mid-transit are found before (or after, respectively) a local minimum. This is the signature of a low-obliquity star[6].

We used both of the above-mentioned methods to establish quantitative bounds on the obliquity: (1) quantifying the relationship between the anomalies and QPV; and (2) modelling a particular pair of transits for which the anomalies can be attributed to transits over the same spot. To support both of these methods, we determined the basic transit parameters—such as the planet-to-star radius ratio (R_{pl}/R_{star}) and impact parameter—by fitting the transit data with a standard model for the loss of light during a planetary transit[12]. We excluded the anomalies from the fit, and accounted for transit depth variations due to unocculted spots (see Supplementary Information, and Supplementary Fig. 1). Results are given in Table 1.

The premise of the first method is that any spot that causes an anomaly must also contribute to the QPV. For a given spatial orientation of the star, geometry dictates a specific relationship between the timing of the anomaly and the phase of the QPV. However, all spots contribute to the QPV, not just the occulted spot. Therefore, to measure the obliquity, we must associate each anomaly with a particular component of the QPV. Out of concern that such associations are ambiguous, we exhaustively tried all plausible associations. We rank-ordered the anomalies in order of the loss of light produced by the spot, and focused attention on the five strongest anomalies. We measured the time of each anomaly relative to mid-transit, as well as the time of the transit relative to each local minimum in the QPV within a rotation period (see Supplementary Information). For one of the anomalies there is only one plausible choice for the associated local minimum, whereas in each of the other four cases there are two candidate local minima, giving a set of 16 possible associations. We find that only one of these 16 is compatible with a single orientation of the host star, and in that case the stellar equator is aligned on the sky with the planet's orbit (see Fig. 1, and Supplementary Fig. 2). We explored all allowed orientations with a Monte Carlo Markov Chain (MCMC) algorithm[13], finding the sky-projected obliquity to be 4° ± 10°.

For the second method, we searched for pairs of anomalies produced by the same spot. Between successive transits of planet c, a spot will rotate 3.77 times around the star, thereby advancing in longitude by 0.77 of a full circle or 277°, relative to the meridian defined by the sky projection of the stellar rotation axis. An advance by 277° is equivalent to regression by 83°. Therefore, if a spot persists for at least four rotations, and if the spot's trajectory is parallel to the planet's trajectory

[1]Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [2]Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. [3]Bay Area Environmental Research Institute, 560 Third Street West, Sonoma, California 95476, USA. [4]NASA Ames Research Center, Moffett Field, California 94035, USA. [5]SETI Institute, 189 Bernardo Ave. no. 100, Mountain View, USA. [6]University of Florida, 211 Bryant Space Science Center, Gainesville, Florida 32611-2055, USA. [7]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [8]Department of Astronomy, University of California, Berkeley, California 94720, USA.

Table 1 | Parameters of the host star Kepler-30, starspots and planets

Host star parameters*

KIC/KOI number	3832472/806
Kepler magnitude	15.4
Mass (M_{Sun})	0.99 ± 0.08
Radius (R_{Sun})	0.95 ± 0.12
Effective temperature (K)	$5{,}498 \pm 54$
Gyrochronology age estimate (Gyr)	2.0 ± 0.8
Quadratic LD coefficient, u_1	0.38 ± 0.09
Quadratic LD coefficient, u_2	0.40 ± 0.19
Linear LD coefficient, u	0.54 ± 0.02
Stellar density (g cm^{-3})	2.00 ± 0.10

Planetary parameters†

Parameter	Kepler-30b	Kepler-30c	Kepler-30d
Orbital period (days)	29.334 ± 0.008	60.3231 ± 0.0002	143.343 ± 0.009
Mid-transit time (BJD)	2455246.65 ± 0.04	2455357.8870 ± 0.0005	2455273.530 ± 0.010
Eccentricity, e (degrees)	0.042 ± 0.003	0.0111 ± 0.0010	0.022 ± 0.005
Periapse angle, ω (degrees)	-31 ± 7	-49 ± 6	-163 ± 7
Nodal angle, Ω (degrees)	0.03 ± 0.17	0 (relative to 30c)	1.3 ± 0.5
Planetary mass (M_E)	11.3 ± 1.4	640 ± 50	23.1 ± 2.7
$\lvert I - 90^\circ \rvert$ (degrees)	0.18 ± 0.16	0.32 ± 0.03	0.16 ± 0.02
Impact parameter	$0.38^{+0.12}_{-0.20}$	$0.40^{+0.04}_{-0.06}$	$0.38^{+0.08}_{-0.14}$
$(R_{pl}/R_{star})^2$	0.00165 ± 0.00008	0.0162 ± 0.0008	0.0083 ± 0.0004
Planet density (g cm^{-3})	1.02 ± 0.13	1.88 ± 0.17	0.19 ± 0.02
Planet radius (R_E)	3.9 ± 0.2	12.3 ± 0.4	8.8 ± 0.5

Starspot parameters and spin-axis orientation‡

Spot rotational period (days)	16.0 ± 0.4
Spot intensity relative to unspotted photosphere	0.85 ± 0.03
Inferred spot temperature (K)	$5{,}298 \pm 65$
Angular radius of spot (degrees)	21^{+7}_{-3}
Sky-projected obliquity, recurrence method (degrees)	-1 ± 10
Sky-projected obliquity, 5-anomaly method (degrees)	4 ± 10

* Most of the host star parameters are obtained from the literature, and are based on the analysis of high-resolution spectra in conjunction with stellar-evolutionary models[8]. The limb darkening (LD) coefficients are obtained from the light curve analysis (see Supplementary Information). The stellar density is obtained from the dynamical modelling of transit timings and durations.

† Most of the planet parameters are obtained from the four-body dynamical model (see Fig. 3, Supplementary Table 4), with the exceptions of the impact parameters and $(R_{pl}/R_{star})^2$, which are obtained strictly from the light curve analysis. Periods and epochs are best-fits to constant-period models, with error bars reflecting the 1 s.d. spread in the transit timing measurements. $\lvert I - 90^\circ \rvert$ is the deviation of the orbital inclination I from 90° (edge-on). The results for the planetary masses and radii take into account the uncertainty in the assumed stellar mass. The results for $(R_{pl}/R_{star})^2$ are assigned a relative error of 5% to account for possible contamination of the Kepler photometric aperture by background stars. The mass and radius of planet c agree with theoretical models of gas giant planets[20] (see Supplementary Information).

‡ The spot parameters are obtained from the spot model (see Fig. 2). In all cases the quoted results and statistical uncertainties are based on the 15.85%, 50% and 84.15% levels of the cumulative a posteriori probability distribution (marginalizing over all other parameters), as determined with the MCMC algorithm.

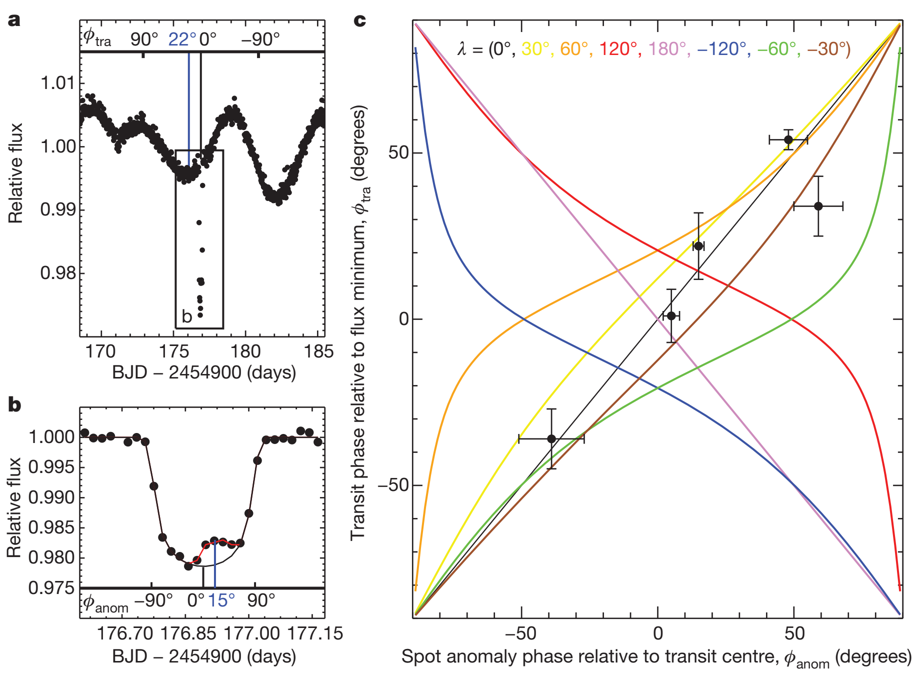


Figure 1 | Evidence for a low obliquity based on transits over several starspots at differing stellar longitudes. **a**, A portion of the Kepler light curve, with a box highlighting the transit of Kepler-30c shown in panel **b**. The transit occurred just after a local minimum in the QPV. The time of the transit is measured with respect to the selected flux minimum, divided by the rotation period and expressed in degrees, giving the 'transit phase', $\phi_{tra} = 22^\circ \pm 10^\circ$. The transit phase is also computed relative to all other local minima within one rotation period. BJD, barycentric Julian day. **b**, A flux anomaly is observed during the transit. The black line is a model without starspots and the red line is a model with one spot. The 'anomaly phase', which can be directly compared to the transit phase, is defined by $\sin \phi_{anom} = 2x/L$, where x is the distance from the spot to the centre of the transit chord, and L is length of the transit chord. In this case $\phi_{anom} = 15^\circ \pm 2^\circ$, in agreement with ϕ_{tra} and consistent with a low obliquity. **c**, Coloured lines show the expected relation between ϕ_{anom} and ϕ_{tra}, for different sky-projected obliquities (λ) and a fixed stellar inclination of 90°. Because the association between anomalies and minima may be ambiguous, ϕ_{tra} was computed for all plausible associations, for the five largest spot anomalies. Only one such set of associations is consistent with a single choice of the stellar orientation. Shown here for that unique choice of associations (see Supplementary Table 3) is the observed relation between ϕ_{anom} and ϕ_{tra} implying a projected obliquity $\lambda = 4^\circ \pm 10^\circ$. This error, and the errors on all phases, is ±1 s.d.

(that is, if the obliquity is low), then an anomaly observed in the second half of a transit should be followed by an anomaly in the first half of the next transit. The two anomalies should differ by 83° in the suitably defined 'anomaly phase' (see Fig. 1).

Two of the five strongest anomalies have this expected phase relationship (see Fig. 2), corroborating the finding of a low obliquity. The QPV produced by this spot is coherent over the interval spanned by the two transits, confirming the persistence of the spot (see Supplementary Table 2). Figure 2 shows a spot model fitted to the transit data. For completeness, three spots were included in the model, although only the largest spot (labelled 1) bears information on the stellar obliquity, because it was transited twice by planet c. The model parameters include the spin orientation of the star, the rotation period, and the spot properties (sizes, locations and intensities). Because the rotation period and spot properties are constrained externally from the QPV, the model could be used to constrain the spin orientation, with results given in Table 1 (see Supplementary Information for details), including a sky-projected obliquity $-1^\circ \pm 10^\circ$. This low sky-projected obliquity is likely to be representative of the true obliquity[14].

Furthermore, all three planetary orbits must be nearly coplanar. The mere existence of multiple transiting planets suggests coplanarity[15], although the possibility remains that the orbits are mutually inclined with nodes (lines of intersection) that happen to lie along the line of sight. However, for Kepler-30, such mutual inclinations would be detectable through variations in transit times and durations caused by nodal precession. To quantify this argument, we performed a four-body integration of Newton's equations[16–18]. To be compatible with the observed transit times and durations, the mutual inclinations must be smaller than a few degrees. A by-product of our dynamical analysis combined with the transit analysis and the known mass of the star[7] is the determination of the planetary masses and radius (Fig. 3, Table 1).

Such an orderly arrangement might seem to be a natural consequence of the standard model of planet formation, based on core accretion within a flat disk[19]. Recently, though, the host stars of many 'hot Jupiter' systems have been found with high obliquities, in some cases even spinning backward relative to the planetary orbit[3,4]. Indeed, it has been argued that stars with hot Jupiters had initially random obliquities, and the only reason low obliquities are more frequent than expected is the obliquity-damping effect of planet–star tidal interactions[4].

The observed high obliquities in hot-Jupiter systems have been interpreted as evidence that hot Jupiters attained their close-in orbits through dynamical interactions (which can strongly perturb a planet's orbital orientation) followed by tidal capture. This view is in opposition to the previous paradigm for the origin of hot Jupiters, in which a gradual

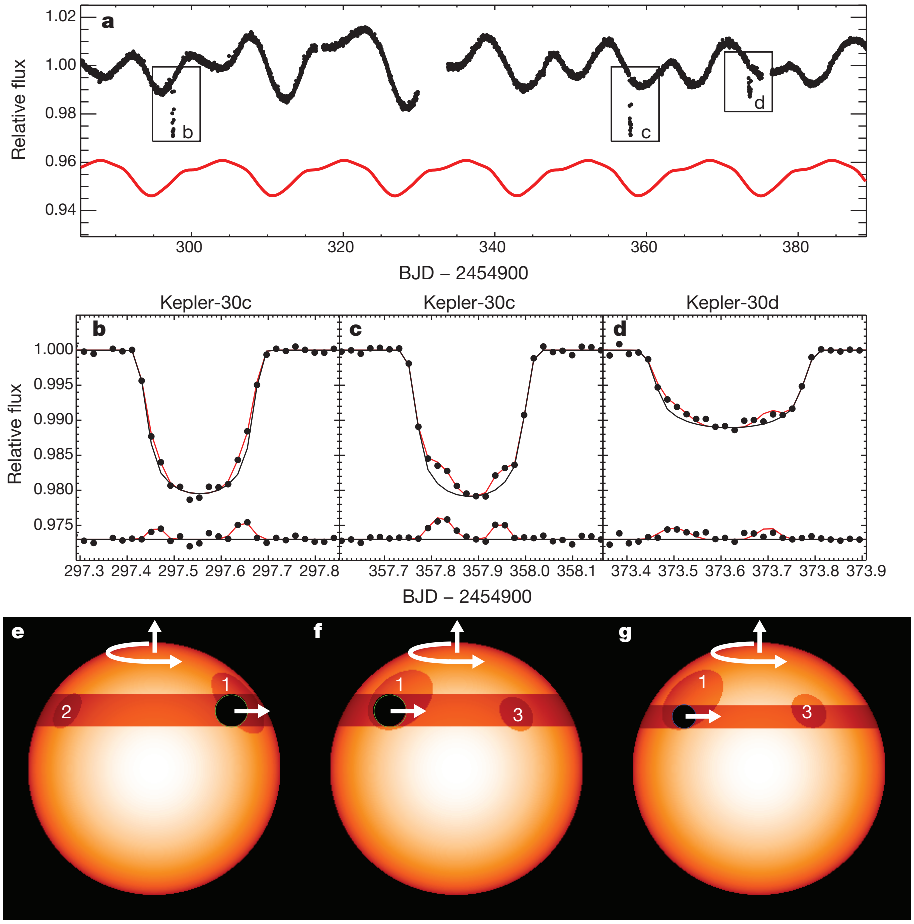


Figure 2 | Evidence for a low obliquity based on a consecutive pair of transits over a single starspot. **a**, Data points (black) are a portion of the Kepler light curve, showing the QPV with an approximate 16-day period. The red curve is a model consisting of three spots (shifted vertically for clarity). The model does not take into account spot evolution or differential rotation and is not expected to fit perfectly. Three particular transits are highlighted with boxes and labelled for subsequent discussion. **b**, Light curve of a transit of planet c. The solid dots are data points, the black curve is a transit model with no spots, and the red curve is the best-fitting model with three spots. Residuals from the best-fitting model are displayed near the bottom of the plot. **c**, Same as the previous panel, but for the next transit of planet c. **d**, Same as the previous panel, but for the next transit of planet d. The key parameter of the model, the projected obliquity, was constrained to be smaller than 10°. **e**, Illustration of the stellar disk, dark spots and transit chord for the time range plotted in panel **b**. The white arrows convey the direction of stellar rotation. The black disk represents the transiting planet. **f**, **g**, Same as panel **e**, but for the time ranges plotted in panels **c**, **d** respectively. Panels **e** and **f** show that spot 1 was twice eclipsed by planet c, with nearly four stellar rotation periods between the transits. Then, one stellar rotation later, spot 1 was also eclipsed by planet d (panel **g**). (The smaller spot, 3, may also have been eclipsed by both planets during this time interval, though the eclipse by planet d is not securely detected.)

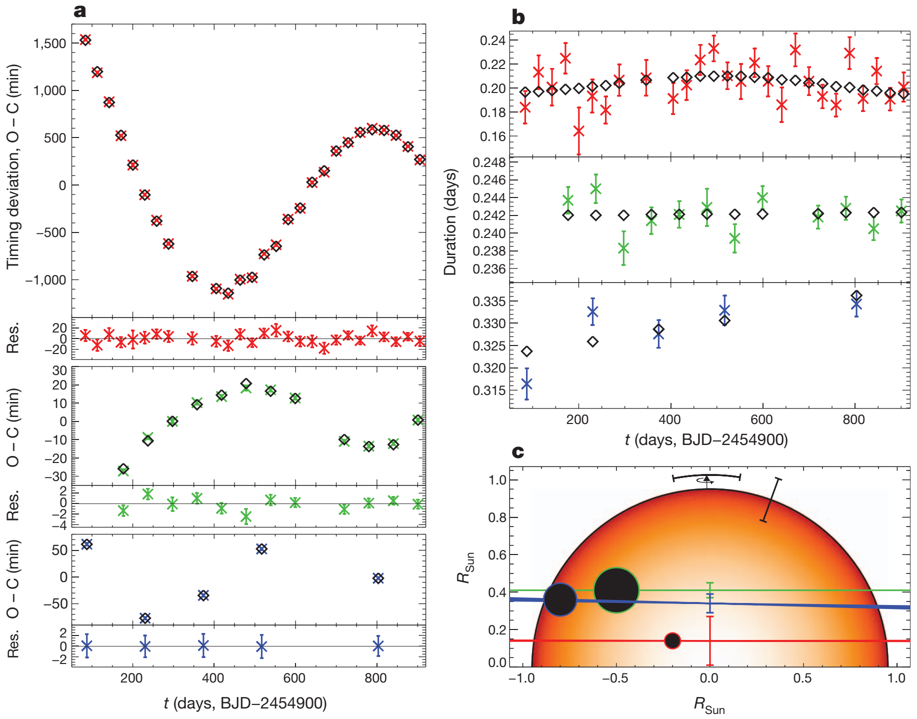


Figure 3 | Evidence for coplanar planetary orbits based on an analysis of transit times and durations. Throughout, sub-panels and diagram components using colours red, green and blue refer to planets Kepler-30b, Kepler-30c and Kepler-30d, respectively. **a**, The deviation of individual observed (O) transit times (see Supplementary Table 1 with all errors defined as ± 1 s.d.) from a constant-period calculation (C) versus time. For planets c and d, suspected starspot-crossing anomalies were masked out before the analysis. Calculated transit times based on a model including planet–planet dynamical interactions[8,18] (Supplementary Information) are shown as open diamonds. Residuals (Res.) between the data and the model are shown below each O − C plot. **b**, Observed and modelled transit durations. Here the transit duration is defined as the length of time when the centre of the planet is projected in front of the stellar disk. **c**, Diagram of the paths of the planets (black circles with coloured rims) across the face of the star. The error bars show the uncertainty in the impact parameters of the orbits, which are constrained from the timescale of ingress and egress. For planets b and d, three lines are shown, delimiting the 1 s.d. region allowed for the rotations around the line of sight, relative to planet c. The lack of secular changes in the durations (**b**) implies coplanarity to within a few degrees. The error bar on the stellar limb (upper right) is the uncertainty in the stellar radius. The stellar spin axis is denoted (upper middle); its projected orientation is determined from the starspot analysis to be aligned with the planets to within 10° (1 s.d.).

transfer of energy and angular momentum to the protoplanetary disk causes their orbits to shrink (and maintain a fixed orientation).

One reason why the scenario involving dynamical and tidal interactions has not gained universal acceptance is that obliquity measurements were previously confined to giant planets with small periastron distances. One would like to make sure that the high obliquities are indeed confined to systems that have experienced dynamical interactions. Otherwise it remains possible that stars and their disks are generally misaligned for reasons unrelated to planets, such as chaotic accretion[1], magnetic interactions[2] or differential torques produced by a neighbouring star.

Kepler-30 is the type of system that needed to be checked: the coplanarity of the planetary orbits suggests a quiescent history without disruptive dynamical interactions, and the planets are too far from the star for strong tidal interactions. The system was selected by virtue of significant spot-crossing anomalies, and not by any criterion that would have biased the result towards low obliquity. Therefore the observed low obliquity is a clue that star–disk misalignments are not the correct explanation for the high obliquities of hot Jupiter hosts, and that hot Jupiters arise from dynamics and tidal capture. There is only a 6% chance of observing such a low obliquity for Kepler-30 if obliquities were drawn from a random initial distribution. To strengthen our interpretation, additional observations of coplanar multiple-planet system are warranted, and are predicted to yield low obliquities.

Received 25 April; accepted 8 June 2012.

1. Bate, M. R., Lodato, G. & Pringle, J. E. Chaotic star formation and the alignment of stellar rotation with disc and planetary orbital axes. *Mon. Not. R. Astron. Soc.* **401,** 1505–1513 (2010).
2. Lai, D., Foucart, F. & Lin, D. N. C. Evolution of spin direction of accreting magnetic protostars and spin-orbit misalignment in exoplanetary systems. *Mon. Not. R. Astron. Soc.* **412,** 2790–2798 (2011).
3. Triaud, A. *et al.* Spin-orbit angle measurements for six southern transiting planets. New insights into the dynamical origins of Hot Jupiters. *Astron. Astrophys.* **524,** A25–A46 (2010).
4. Winn, J. N., Fabrycky, D., Albrecht, S. & Johnson, J. A. Hot stars with hot Jupiters have high obliquities. *Astrophys. J.* **718,** L145–L149 (2010).
5. Sanchis-Ojeda, R. *et al.* Starspots and spin-orbit alignment in the WASP-4 exoplanetary system. *Astrophys. J.* **733,** 127–135 (2011).
6. Nutzman, P. A., Fabrycky, D. C. & Fortney, J. J. Using star spots to measure the spin-orbit alignment of transiting planets. *Astrophys. J.* **740,** L10–L14 (2011).
7. Désert, J. M. *et al.* The hot-Jupiter Kepler-17b: discovery, obliquity from stroboscopic starspots, and atmospheric characterization. *Astrophys. J.* **197** (Suppl.), 14–26 (2011).
8. Fabrycky, D. C. *et al.* Transit timing observations from Kepler: IV. Confirmation of 4 multiple planet systems by simple physical models. *Astrophys. J.* **750,** 114–130 (2012).
9. Winn, J. N. *et al.* Measurement of spin-orbit alignment in an extrasolar planetary system. *Astrophys. J.* **631,** 1215–1226 (2005).
10. Borucki, W. J. *et al.* Kepler planet-detection mission: introduction and first results. *Science* **327,** 977–980 (2010).
11. Scargle, J. D. Studies in astronomical time series analysis. II — Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263,** 835–853 (1982).
12. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580,** L171–L175 (2002).
13. Ford, E. Improving the efficiency of Markov Chain Monte Carlo for analyzing the orbits of extrasolar planets. *Astrophys. J.* **642,** 505–522 (2006).
14. Fabrycky, D. C. & Winn, J. N. Exoplanetary spin-orbit alignment: results from the ensemble of Rossiter-McLaughlin observations. *Astrophys. J.* **696,** 1230–1240 (2009).
15. Lissauer, J. *et al.* Architecture and dynamics of Kepler's candidate multiple transiting planet systems. *Astrophys. J.* **197** (Suppl.), 8–33 (2011).
16. Holman, M. J. & Murray, N. W. The use of transit timing to detect terrestrial-mass extrasolar planets. *Science* **307,** 1288–1291 (2005).
17. Agol, E., Steffen, J., Sari, R. & Clarkson, W. On detecting terrestrial planets with timing of giant planet transits. *Mon. Not. R. Astron. Soc.* **359,** 567–579 (2005).

18. Holman, M. J. *et al.* Kepler-9: a system of multiple planets transiting a sun-like star, confirmed by timing variations. *Science* **330,** 51–54 (2010).
19. Lissauer, J. J. Planet formation. *Annu. Rev. Astron. Astrophys.* **31,** 129–174 (1993).
20. Fortney, J. J., Marley, M. S. & Barnes, J. W. Planetary radii across five orders of magnitude in mass and stellar insolation: application to transits. *Astrophys. J.* **659,** 1661–1672 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Kepler was competitively selected as the tenth Discovery mission. Funding for this mission was provided by NASA's Science Mission Directorate. The data presented in this Letter were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts. S. Albrecht, E. Agol, J. A. Carter, L. Doyle and A. Shporer provided comments on the manuscript. D.C.F. acknowledges NASA support through Hubble Fellowship grant HF-51272.01-A, awarded by STScI. D.R. acknowledges the Harvard Institute for Theory and Computation. E.B.F., M.J.H. and J.N.W. acknowledge NASA support through the Kepler Participating Scientist programme.

Author Contributions R.S.-O. performed the spot analysis and wrote the first draft of the paper. D.C.F. performed the dynamical analysis, contributed to the spot modelling, and helped write the paper. J.N.W. contributed to the modelling and helped write the paper. The remaining authors listed below contributed equally. T.B., B.D.C., F.M., S.E.S., M.S. and S.E.T. worked on the data collection, processing and review that yielded the time-series photometry. E.B.F. and M.J.H. provided feedback on the text and interpretation. J.J.F. and J.J.L. worked on elucidating the structure and radii of the planets. J.C.G. worked on the development of Kepler spacecraft photometer electronics, is a builder of Keplercam for the Kepler Input Catalog and follow-up spectral typing. A.W.H. contributed the Keck-HIRES spectra from which the stellar properties were derived. J.M.J. is the Co-Investigator for Data Analysis and designed and built the pipeline that produced the light curves on which this paper is based. D.K. contributed to the concept, design, development, testing and commissioning of the Kepler Mission. G.W.M. contributed the Keck spectroscopy and helped with the imaging, and with some other parts of the original follow-up observations. D.R. studied the possibility of mutual transits and provided feedback on the technical details of the analysis. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.S.-O. (rsanchis86@gmail.com) or D.C.F. (daniel.fabrycky@gmail.com).

doi:10.1038/nature11314

Oligodendroglia metabolically support axons and contribute to neurodegeneration

Youngjin Lee[1]*, Brett M. Morrison[1]*, Yun Li[1], Sylvain Lengacher[2], Mohamed H. Farah[1], Paul N. Hoffman[1], Yiting Liu[1], Akivaga Tsingalia[1], Lin Jin[1], Ping-Wu Zhang[1], Luc Pellerin[3], Pierre J. Magistretti[2] & Jeffrey D. Rothstein[1,4,5]

Oligodendroglia support axon survival and function through mechanisms independent of myelination, and their dysfunction leads to axon degeneration in several diseases. The cause of this degeneration has not been determined, but lack of energy metabolites such as glucose or lactate has been proposed. Lactate is transported exclusively by monocarboxylate transporters, and changes to these transporters alter lactate production and use. Here we show that the most abundant lactate transporter in the central nervous system, monocarboxylate transporter 1 (MCT1, also known as SLC16A1), is highly enriched within oligodendroglia and that disruption of this transporter produces axon damage and neuron loss in animal and cell culture models. In addition, this same transporter is reduced in patients with, and in mouse models of, amyotrophic lateral sclerosis, suggesting a role for oligodendroglial MCT1 in pathogenesis. The role of oligodendroglia in axon function and neuron survival has been elusive; this study defines a new fundamental mechanism by which oligodendroglia support neurons and axons.

Oligodendroglia promote rapid conduction of action potentials by ensheathing central nervous system (CNS) axons with myelin. Oligodendrocyte diseases, such as multiple sclerosis and leukodystrophies, have demonstrated demyelination and axon degeneration at autopsy[1,2]. Mouse models of oligodendrocyte injury, including proteolipid protein 1 (*Plp1*)-null mice[3] and *Cnp* (also known as *Cnp1* and *CNPase*) mutant mice[4], demonstrate axon loss without considerable demyelination, suggesting that oligodendroglia support axon survival through a myelin-independent mechanism, possibly as a result of insufficient axonal energy support[5]. Myelinated axons are only exposed to extracellular energy substrates at the nodes of Ranvier, and therefore may require specialized transport of energy metabolites from myelinating oligodendroglia to meet their high metabolic needs. The identity of these metabolites is unclear, but our study suggests that lactate may be essential and its transport dependent on MCT1 (Supplementary Fig. 1).

MCT1, along with the transporters MCT2 and MCT4, transport monocarboxylic acids (that is, lactate, pyruvate and ketone bodies), and localize to the CNS[6]. Neurons express MCT2, and glia express both MCT1 and MCT4 (refs 7, 8), although MCT1 is the dominant glial transporter in the brain[9]. Recently, MCT1 was localized to oligodendroglia and MCT2 to axons of the corpus callosum and cerebellar white matter by immunohistochemistry[10]. *In vitro* astrocytes produce lactate through aerobic glycolysis[11,12], and lactate alone can support neurons in the absence of glucose, presumably through MCT2 localized to neurons. This hypothetical energy transfer was termed the astrocyte–neuron lactate shuttle[13]. Support for lactate-based neuronal support has come from both *in vitro* and *in vivo* models[14,15]; however, the physiological role for lactate in the non-stressed, uninjured CNS is largely unknown. We now report that oligodendroglia are an important site of MCT1 expression in the brain and spinal cord and are the principal metabolic supplier of lactate to axons and neurons.

Oligodendroglia injury is well established in demyelinating diseases[16], but the supply of energy metabolites to axons could also be crucial in other neurological diseases. In this study, we investigated amyotrophic lateral sclerosis (ALS), a fatal neurological disease characterized clinically by progressive weakness and pathologically by cortical and spinal motoneuron degeneration. Although the pathogenesis of motoneuron degeneration is unknown, it is mediated partly by surrounding astroglia and microglia[17]. A recent study suggests that grey matter oligodendroglia may be injured in ALS[18], and we propose that reduced expression of MCT1 is one mechanism by which oligodendroglia produce neurotoxicity in ALS.

MCT1 localizes to oligodendroglia *in vivo*

Astrocytes[8,19,20], ependymocytes, endothelial cells[19,21] and recently oligodendroglia[10] have inconsistently been demonstrated to express MCT1. This variability is due to limitations in antibody specificity and/or affinity, along with differences in the species and age of samples analysed. To overcome this technical challenge, we produced two lines of bacterial artificial chromosome (BAC) transgenic mice carrying the tdTomato fluorescent reporter for cellular localization and *in vivo* expression level of *MCT1* messenger RNA in the CNS and peripheral organs (Fig. 1 and Supplementary Figs 2–4). Results are shown for the highest expressing line, although cellular localization was identical for the second line (data not shown). Enrichment of *MCT1* mRNA was found within fluorescence-activated cell sorted (FACS) tdTomato-positive cells (Fig. 1a, group ii), verifying the specificity of the reporter. Expression was similar in perinatal mice, although reporter expression around blood vessels was increased (data not shown). MCT1 BAC mice were crossed with MOBP–eGFP and GLT1–eGFP BAC reporter mice[22], which express enhanced green fluorescent protein (eGFP) driven by the oligodendrocyte-specific myelin-associated oligodendrocyte protein (MOBP) and astrocyte-specific glutamate transporter 1 (GLT1, also known as

[1]Department of Neurology, The Johns Hopkins University, 855 North Wolfe Street, Rangos 248, Baltimore, Maryland 21205, USA. [2]School of Life Sciences, Brain Mind Institute, Ecole Polytechnique Federale de Lausanne (EPFL), Station 19, CH-1015 Lausanne, Switzerland. [3]Department of Physiology, University of Lausanne, 7 Rue du Bugnon, CH-1005 Lausanne, Switzerland. [4]Department of Neuroscience, The Johns Hopkins University, 855 North Wolfe Street, Rangos 270, Baltimore, Maryland 21205, USA. [5]The Brain Science Institute, The Johns Hopkins University, 855 North Wolfe Street, Rangos 270, Baltimore, Maryland 21205, USA.
*These authors contributed equally to this work.

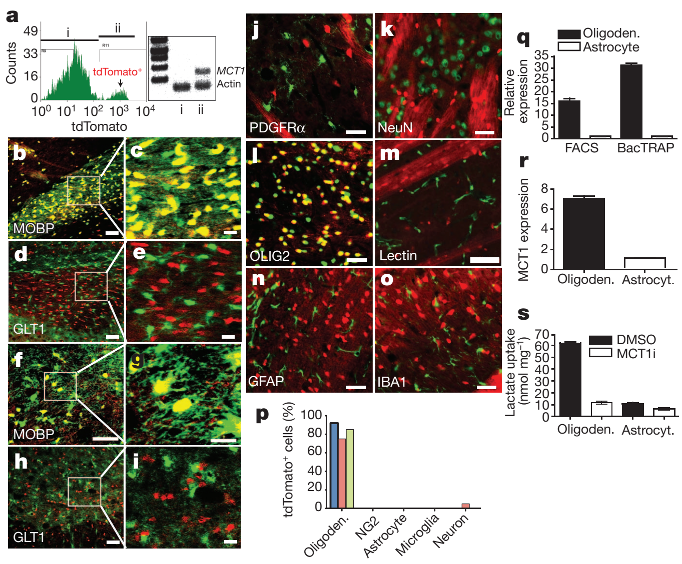


Figure 1 | MCT1 expressed primarily within oligodendroglia in the CNS. **a**, Histogram and rtPCR for *MCT1* mRNA from FACS-sorted cells with (ii) or without (i) tdTomato expression. **b–e**, Corpus callosum of MCT1–tdTomato reporter mice crossed with MOBP–eGFP (**b**, **c**) or GLT1–eGFP (**d**, **e**) BAC mice. **f–i**, Spinal cord of MCT1– tdTomato reporter mice crossed with MOBP–eGFP (**f**, **g**) or GLT1–eGFP (**h**, **i**) BAC mice. Scale bars, 50 μm (low power, left) and 14 μm (high power, right). **j–o**, MCT1–tdTomato reporter mouse brain immunostained with cell-specific markers. Scale bars, 12.5 μm (except 50 μm for lectin). **p**, Percentage of tdTomato cells co-labelled with CNS cell markers in the spinal cord (blue), cortex (red) and corpus callosum (green). **q**, **r**, rtPCR of *MCT1* mRNA isolated from oligodendroglia and astroglia by FACS and BacTRAP techniques (**q**) and human oligodendroglioma (M03.13) and astrocytoma (U87) cell lines (**r**; means of four replicates). **s**, Lactate transport and blockade by selective MCT1 inhibitor (MCT1i) in oligodendroglioma and astrocytoma ($n = 4$). DMSO, dimethylsulphoxide. Error bars denote s.e.m.

SLC1A2), respectively. *MCT1* mRNA was almost exclusively localized to oligodendroglia in the brain and spinal cord (Fig. 1 and Supplementary Fig. 3), with greater than 70–80% co-localization in the spinal cord, cortex and corpus callosum (Fig. 1p). Rare neuronal populations expressed MCT1 (Supplementary Fig. 4), although none in retinal ganglion cells (Supplementary Fig. 2k) or spinal cord motoneurons (Supplementary Fig. 4o, p). Surprisingly, there was virtually no expression of *MCT1* mRNA within adult CNS astrocytes (Fig. 1 and Supplementary Fig. 3), nor was it found in NG2 cells, endothelial cells or microglia (Fig. 1). Most MCT1 BAC-positive cells co-labelled with an oligodendroglia lineage marker, oligodendrocyte transcription factor 2 (OLIG2; Fig. 1l)[23], and few other CNS cells expressed tdTomato (Fig. 1p).

MCT1 protein also co-localized with myelinating oligodendroglia, as MCT1 immunoreactivity co-localized with myelin basic protein (MBP) and the oligodendrocyte-specific marker 2′,3′-cyclic-nucleotide 3′-phosphodiesterase (CNP) in rodent and human brain (Supplementary Fig. 5). MCT1 protein was closely aligned with axons (Supplementary Fig. 5h–j), but not astroglia, axonal nodes or paranodes (Supplementary Fig. 5k–p). *In vitro*, NG2 cells and primary oligodendroglia both expressed the MCT1–tdTomato reporter (Supplementary Fig. 6a–f), as do primary astrocyte cultures (Supplementary Fig. 6g–l)[24], which may account for the discrepancies in MCT1 localization in the literature.

The cell specificity and enrichment of native *MCT1* mRNA was further evaluated by FACS and BAC translating ribosome affinity purification (BacTRAP) techniques. *MCT1* mRNA, quantified by real-time PCR with reverse transcription (real-time RT–PCR) of acutely isolated mouse brain mRNA from FACS and BacTRAP astrocytes and oligodendroglia, was expressed at 16-fold (FACS) and 31-fold (BacTRAP) higher levels in mature oligodendroglia than in astrocytes (Fig. 1q). MCT1 expression (Fig. 1r) and functional lactate transport (Fig. 1s) were also detected and enriched in oligodendrocyte tumour cells (that is, oligodendroglioma) as compared with astrocyte tumour cells (that is, astrocytoma). Although not detectable by either the MCT1–tdTomato BAC reporter mice or immunohistochemistry, FACS and BacTRAP experiments suggest that astrocytes express very low levels of *MCT1* mRNA *in vivo*; although it is clear that MCT1 expression and lactate transport is much greater in oligodendroglia than astrocytes. Astrocytes *in vivo* express the MCT4 lactate transporter (Yo.L., unpublished observations)[25,26], suggesting that this transporter may have a more important role in astrocytes than MCT1.

Neuronal survival *in vitro* requires MCT1

Similar to *in vivo*, *MCT1* mRNA is primarily localized to oligodendrocyte-lineage cells, immunostained with OLIG2, in postnatal organotypic spinal cord cultures (Supplementary Fig. 7). To investigate whether downregulation or inhibition of MCT1 produces neuron death, organotypic spinal cord cultures were treated with antisense oligonucleotides (ASO) or a specific MCT1 transport inhibitor (MCT1i; Supplementary Fig. 7i)[27]. After three weeks of treatment, only 67% ± 6.7 (mean ± s.e.m.) of motoneurons survived in the ASO group ($P < 0.01$, Fig. 2a, b), and 66% ± 5.9 in the MCT1i group ($P < 0.05$, Fig. 2c). Thus, genetic or pharmacological reduction of MCT1 led to motoneuron death.

If the function of MCT1 is to export lactate from oligodendroglia and provide energy metabolites to neurons, then neurons exposed to glucose-free media or stimuli that increase neuronal activity should be particularly vulnerable. Motoneurons in organotypic spinal cord slice cultures were resistant to short periods (2 h) of glucose deprivation,

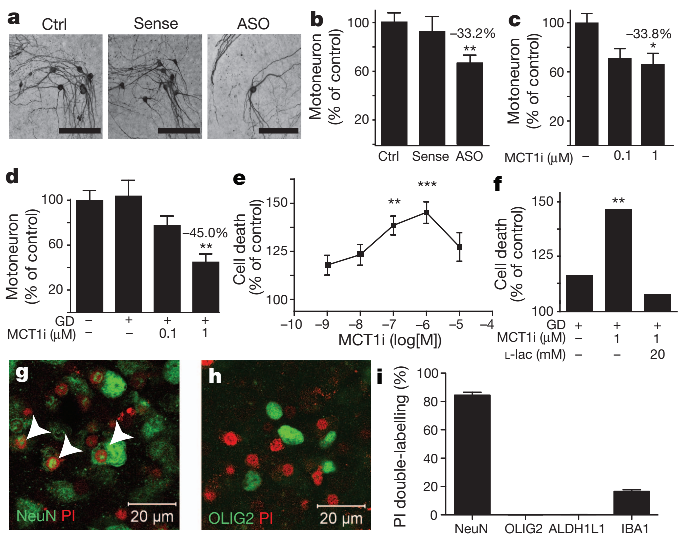


Figure 2 | MCT1 required for neuronal survival *in vitro*. **a–d**, Photomicrographs (**a**) and quantification (**b–d**) of motoneurons in spinal cord slice cultures treated with media only (ctrl; $n = 105$), MCT1 sense oligonucleotides ($n = 55$), or MCT1 ASO ($n = 101$) for 3 weeks (**b**), after 3 weeks of treatment with MCT1i (**c**; $n = 68$, 30 and 26 sections for columns 1–3, respectively), or 2 h of glucose deprivation (GD) with or without MCT1i (**d**; $n = 83$, 90, 68 and 70 sections for columns 1–4, respectively). **e**, **f**, Propidium iodide uptake in slice cultures treated with 2 h glucose deprivation plus MCT1i (**e**; $n = 79$, 14, 14, 39, 45 and 10 sections for control, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6} and 10^{-5} MCT1i, respectively) or 2 h glucose deprivation with or without MCT1i and 20 mM lactate (**f**; $n = 15$ for all groups). **g**, **h**, Propidium iodide uptake in slices treated with glucose deprivation and MCT1i, labelled with a neuronal (**g**; NeuN, co-localized cells marked with arrowheads) or an oligodendroglia (**h**; OLIG2) marker. Scale bars, 20 μm. **i**, Percentage of propidium iodide-labelled cells co-localizing with cell-specific markers ($n = 11$, 10, 10 and 11 sections for NeuN, OLIG2, ALDH1L1 (an astrocyte marker) and IBA1, respectively). Error bars denote s.e.m. $*P < 0.05$; $**P < 0.01$.

presumably owing to the presence of glycogen stores in astrocytes[28]. Simultaneous exposure to glucose deprivation and MCT1i led to concentration-dependent motoneuron loss, as measured by counts of neurofilament-containing neurons in the ventral horn ($P < 0.01$, Fig. 2d), and overall cell death, as measured by propidium iodide uptake ($P < 0.001$, Fig. 2e). Propidium iodide uptake occurred mainly within neurons ($P < 0.001$, Fig. 2g, i and Supplementary Fig. 7) and occasionally within IBA1-positive microglia (Fig. 2i and Supplementary Fig. 7). Much of the propidium iodide uptake in microglia is probably due to phagocytosing dead cells; however, limited microglia death cannot be excluded. Importantly, there was no death of oligodendrocyte-lineage cells or astrocytes after glucose deprivation and treatment with MCT1i (Fig. 2h, i and Supplementary Fig. 7). Other neurons were also vulnerable to glucose deprivation and MCT1i treatment, as considerable propidium iodide uptake was seen in the dorsal horn of the spinal cord and in organotypic cultures from neocortex (Supplementary Fig. 7). MCT1i- and glucose deprivation-induced neurodegeneration was completely prevented by supplementing media with 20 mM L-lactate ($P < 0.01$, Fig. 2f), confirming that the toxicity in cultures was due to reduced lactate release from oligodendroglia, not blockage of lactate uptake on neurons or oligodendroglia. Cell death was also produced by co-treatment of organotypic spinal cord cultures with MCT1i and either glutamate or the GABA (γ-aminobutyric acid) antagonist, bicuculline, both of which depolarize neurons in slice cultures ($P < 0.05$, Supplementary Fig. 8)[29,30]. Because glutamate is also a substrate of metabolism, we confirmed that cell death was dependent on neuron depolarization by blocking cell death with the AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and kainate receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) (Supplementary Fig. 8). In summary, neurons *in vitro* are vulnerable to inhibitors of MCT1 lactate transport. Toxicity is potentiated by removing glucose or increasing the metabolic activity of neurons, and is prevented by supplying exogenous lactate.

Downregulation of MCT1 *in vivo* is motoneuron toxic

To downregulate MCT1 *in vivo*, we produced a lentivirus that expressed GFP and *MCT1* short hairpin RNA (shRNA) driven by the cytomegalovirus (CMV) and H1 promoters, respectively (lenti-shRNA). Lenti-shRNA downregulated MCT1 protein both in cultured cells and cervical spinal cord *in vivo* (Fig. 3a). The amount of downregulation *in vivo* is underestimated by western blot because only a fraction of oligodendroglia are transfected by virus. Controls included injection of a lentivirus expressing GFP only (lenti-GFP) and contralateral injections of virus media. Both of the lentiviruses (lenti-shRNA and lenti-GFP) effectively transduced oligodendroglia and astrocytes (Fig. 3b, c). Four weeks after cervical spinal cord injection of lenti-shRNA, 56% ± 5.7 ($n = 10$, $P < 0.05$) of motoneurons near the injection site survived relative to contralateral media injection, whereas injection of lenti-GFP had no significant effect (Fig. 3, $P > 0.05$). In addition, numerous pathological axonal swellings, immunoreactive for the neurofilament marker SMI-32, resulted from lenti-shRNA treatment (Fig. 3g). As expected, motoneuron death in lenti-shRNA-injected spinal cords produced microglial activation, which was not seen using the lenti-GFP control (Fig. 3h, i). Thus, subacute downregulation of MCT1 focally in the spinal cord is sufficient to produce motoneuron death.

Heterozygous *MCT1*-null mice develop CNS axonopathy

MCT1 shRNA treatment was neurotoxic, and we postulated that reducing expression of *MCT1* in a null mouse would also produce neurodegeneration. Complete absence of MCT1 is embryonically lethal, but heterozygous-null mice ($MCT1^{+/-}$), with ~50% reduction in MCT1, breed and mature normally with no gross phenotypic abnormalities. The mice eventually developed an axonopathy in the brain and spinal cord by 8 months of age, with axon swellings

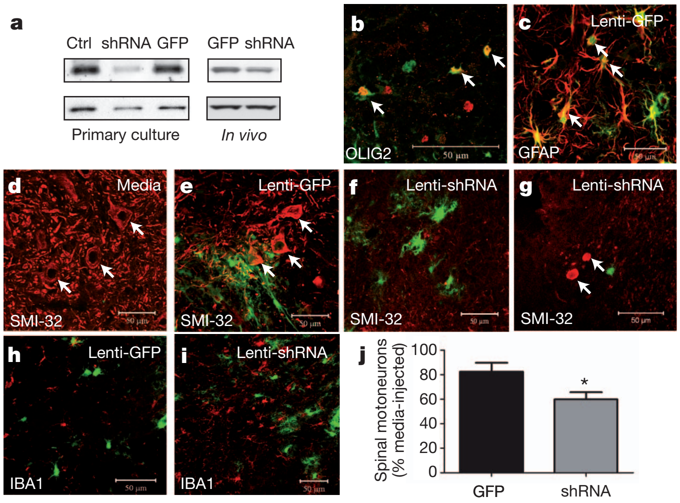


Figure 3 | Lentiviral *MCT1* shRNA is toxic to motoneurons. a, Lenti-shRNA (shRNA) downregulates MCT1 in primary astrocyte cultures and *in vivo* compared with untreated (ctrl) or lenti-GFP control (GFP). **b**, **c**, OLIG2 (**b**) and glial fibrillary acidic protein (GFAP; **c**) in lenti-GFP-injected spinal cords. **d**–**f**, Non-phosphorylated neurofilament (SMI-32) (red) in spinal cords injected with media (**d**), lenti-GFP (**e**), or lenti-shRNA (**f**). Arrows indicate motoneurons. **g**, SMI-32 labels aberrant axonal swellings in lenti-shRNA-injected spinal cords. **h**, **i**, IBA1-positive microglia in spinal cords injected with lenti-GFP (**h**) and lenti-shRNA (**i**). Scale bars, 50 μm. **j**, Quantification of cervical spinal motoneurons injected with lenti-GFP ($n = 8$) or lenti-shRNA ($n = 10$), relative to contralateral media injection. Error bars denote s.e.m. $^*P < 0.05$.

visualized by both light and electron microscopy (Fig. 4 and Supplementary Fig. 9). Notably, axon pathology in these mice is similar to *Cnp*-null mice[31], SOD1 transgenic mice as a model of ALS[32], and to patients with ALS[33]. Despite the axonopathy and activation of astrocytes, microglia and NG2-positive cells (Fig. 4 and

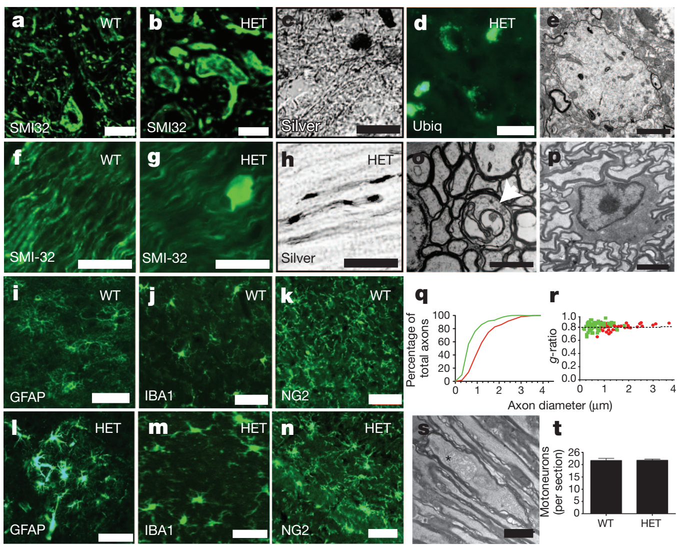


Figure 4 | Heterozygous *MCT1*-null mice develop widespread axonopathy. a–**d**, Spinal cords of wild-type (WT) and $MCT1^{+/-}$ (HET) mice, immunostained for non-phosphorylated neurofilament (SMI-32) (**a**, **b**), silver-stained (**c**), or immunostained for ubiquitin (**d**). **e**, Electron microscopy of axonal spheroid in spinal cord. **f**–**h**, Optic nerves of wild-type and $MCT1^{+/-}$ mice immunostained for SMI-32 (**f**, **g**) or silver-stained (**h**). **i**–**n**, GFAP (**i**, **l**), IBA1 (**j**, **m**), and NG2 (**k**, **n**) in wild-type (**i**–**k**) and $MCT1^{+/-}$ (**l**–**n**) mice. **o**, **p**, Degenerating axon (**o**; arrow) and intact oligodendrocyte (**p**) in optic nerve. **q**, **r**, Axon diameter (**q**) and *g*-ratio (**r**) in $MCT1^{+/-}$ (green line/dots; $n = 158$) and wild-type (red line/dots; $n = 78$) mice. **s**, Distended mitochondrion (asterisk) in $MCT1^{+/-}$ optic nerve. **t**, Number of spinal motoneurons per section from wild-type and $MCT1^{+/-}$ mice ($n = 7$ for each group). Scale bars, 20 μm (**a**–**d**, **f**–**h**), 2 μm (**e**, **o**, **s**), 50 μm (**i**–**n**) and 6 μm (**p**). Error bars denote s.e.m.

Supplementary Fig. 9), there was no change in overall myelination in *MCT1*$^{+/-}$ mice (Supplementary Fig. 9a–f), suggesting that the pathology was not secondary to oligodendrocyte injury.

In the optic nerve, a pure sample of CNS axons the function of which is partly dependent on lactate[34], large and intermediate sized degenerating axons were observed in *MCT1*$^{+/-}$ mice (Fig. 4o). Overall, 1.74 ± 0.06% of optic nerve axons from *MCT1*$^{+/-}$ mice showed morphologic features of degeneration, including swellings, enlarged mitochondria and reduced axon diameter, which were not seen in age-matched littermate control mice (Fig. 4 and Supplementary Fig. 9). Notably, oligodendrocyte morphology (Fig. 4p) and number were not changed, nor was myelination as represented by normal *g*-ratios (the ratio between the diameter of the inner axon and the total out diameter) (Fig. 4r), and intact myelin in non-degenerating axons (Supplementary Fig. 9), again suggesting that axonal degeneration was not due to oligodendrocyte damage or demyelination. Thus, despite only having a partial reduction in MCT1 expression (Supplementary Fig. 9j), adult *MCT1*$^{+/-}$ mice demonstrate widespread CNS axonopathy. Given the low percentage of axons affected, it is not surprising that this axonopathy did not produce a behavioural phenotype, loss of motoneurons (Fig. 4s), or obvious injury to retinal ganglion cells (Supplementary Fig. 9t, u) in *MCT1*$^{+/-}$ mice. In summary, *MCT1*$^{+/-}$ mice develop axon degeneration in the CNS without demyelination or oligodendrocyte injury, suggesting that MCT1 is crucial for the normal function of CNS axons through a myelin-independent mechanism.

Oligodendrocyte-specific MCT1 loss causes axonopathy

Although MCT1 is primarily expressed within oligodendroglia in the CNS, it is not exclusive to oligodendroglia and thus we cannot exclude the possibility that the axon degeneration seen in *MCT1*$^{+/-}$ mice is due to downregulating MCT1 in other cell types. To investigate this, we produced two lentiviral constructs that downregulate MCT1 selectively in oligodendroglia through different mechanisms. First, lentiviral constructs were produced that expressed either the same *MCT1* shRNA used in Fig. 3 plus GFP (lenti-MBP-shRNA) or GFP alone (lenti-MBP-GFP) driven by the MBP promoter. Lenti-MBP-shRNA downregulated MCT1 protein in the optic nerve (Fig. 5a, b) despite transfecting only a small segment of nerve (Fig. 5c, d). Transfection was selective for oligodendrocyte-lineage cells, as seen by co-localization of all GFP-labelled cells with OLIG2 (Fig. 5e, f, arrows) and not astroglia (Fig. 5c, d). At the injection site, there were a few pathologically enlarged SMI-32-positive axons in lenti-MBP-GFP-injected optic nerves owing to nerve trauma (Fig. 5g, arrowheads), but markedly more in optic nerves injected with lenti-MBP-shRNA (Fig. 5h, arrowheads). Distal to the injection, there were increased degenerating axons, both myelin ovoids (arrows) and 'dark axons' (asterisks), in optic nerve injected with lenti-MBP-shRNA compared with lenti-MBP-GFP (Fig. 5i–l). Restricted penetration of the lentivirus into the nerve led to some variability in axon degeneration. Nevertheless, there was a significant increase in the number of degenerating fibres in optic nerves injected with lenti-MBP-shRNA compared with lenti-MBP-GFP ($P < 0.05$, $n = 4$ nerves, Fig. 5m, n).

We also developed a tamoxifen-inducible Cre-dependent lentivirus expressing *MCT1* shRNA, and produced oligodendrocyte-specific *MCT1* knockdown by injecting virus into the corpus callosum of transgenic mice expressing a tamoxifen-inducible Cre (CreER) under the transcriptional control of myelin proteolipid protein (PLP1) (Supplementary Fig. 10a)[35,36]. Viral transfection of corpus callosum oligodendroglia was confirmed in control mice by co-localization of GFP with both CC-1—an antibody against adenomatous polyposis coli that labels oligodendrocytes—and MCT1 reporter (Supplementary Fig. 10b–g; arrowheads). In tamoxifen-treated PLP1-CreER mice, the GFP reporter was excised in oligodendroglia, so specific viral localization of GFP to oligodendroglia is generally not visualized. There was clear axonal degeneration near the injection site in the

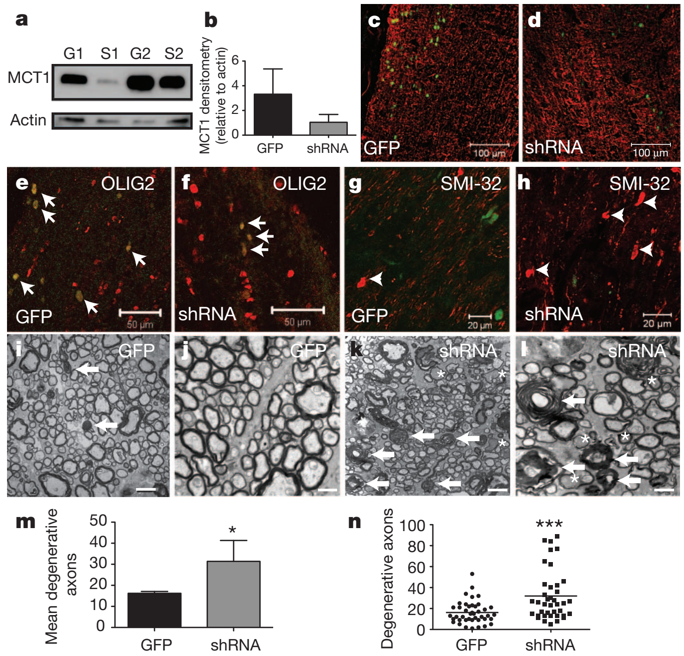


Figure 5 | Selective downregulation of MCT1 in oligodendroglia produces axonal injury. **a**, **b**, Western blot of optic nerves injected with lentivirus expressing MBP-shRNA (S) ($n = 2$) or MBP-GFP (G) ($n = 2$). **c**, **d**, GFAP immunoreactivity in nerve transfected with lenti-MBP-GFP and lenti-MBP-shRNA. **e**, **f**, GFP and shRNA viruses co-localize exclusively with OLIG2 (arrows). **g**, **h**, Axon swellings (arrowheads) labelled with non-phosphorylated neurofilament (SMI-32) after injection of GFP (**g**) and shRNA (**h**) viruses. **i**–**l**, Low-power (**i**, **k**) and high-power (**j**, **l**) magnification electron microscopy photomicrographs after injection of GFP and shRNA viruses. Arrows indicate myelin ovoids and asterisks indicate dark degenerating axons. **m**, **n**, Mean number (**m**; $n = 4$ nerves per group) and scatterplot (**n**; $n = 40$ fields per group) of degenerative axons per electron microscopy field in lenti-GFP- and lenti-shRNA-injected optic nerves. Scale bars, 100 μm (**c**, **d**), 50 μm (**e**, **f**), 20 μm (**g**, **h**), 1 μm (**i**, **k**) and 0.5 μm (**j**, **l**). Error bars denote s.e.m. *$P < 0.05$; ***$P < 0.001$.

PLP1-CreER mice, as demonstrated by SMI-32-immunoreactive axon swellings (Supplementary Fig. 14k–m, boxed areas) and by axonal beading in GFP-labelled axons (Supplementary Fig. 14n), which was not observed when the identical virus was injected into wild-type mice (Supplementary Fig. 14h–j). Thus, *in vivo* oligodendrocyte-specific downregulation of MCT1 in both the optic nerve and the corpus callosum was capable of producing axon degeneration. Overall, these studies confirm that oligodendrocyte MCT1 is crucial for axon survival as loss of lactate transport from oligodendroglia to axons causes axon degeneration in both viral knockdown experiments and *MCT1*$^{+/-}$ mice.

MCT1 expression is reduced in ALS

Recent studies suggest oligodendroglia or their precursors may be injured in mutant SOD1 transgenic mice[18]. We proposed that the reduced ability of grey and white matter oligodendroglia to support motoneurons, caused by altered MCT1 expression, may contribute to ALS pathogenesis as seen in our *in vitro* and *in vivo* studies described earlier. We investigated the expression levels of MCT proteins and the MCT-associated protein CD147 (also known as BSG or basigin) in affected (that is, motor cortex) and unaffected (that is, frontal cortex) regions from patients with ALS and control patients. In addition, we investigated MCT1 reporter activity in the spinal cord of SOD1(G93A) transgenic mice—a commonly used model of ALS that had been crossed with MCT1–tdTomato BAC reporter mice. The motor cortex of patients with ALS showed a greater than 50% decline in MCT1 and MCT4 expression ($P < 0.01$ and $P < 0.001$, respectively) compared with gender- and age-matched control patients

(Fig. 6 and Supplementary Table 1). This reduction was not seen in the unaffected frontal cortex, nor was it seen for CD147, an accessory glycoprotein that increases the localization of MCT1 and MCT4 to the plasma membrane[37] (full blots available in Supplementary Figs 12 and 13). Oligodendroglia were still present in the cortical samples as there was no significant alteration in CNP (Fig. 6a, b), although it is possible the oligodendroglia were immature.

In addition to patients with ALS, downregulation of *MCT1* mRNA was also seen in the spinal cords of early symptomatic (Supplementary Fig. 11) and end-stage (Fig. 6) SOD1(G93A) transgenic mice. SOD1(G93A) transgenic mice were mated to MCT1 BAC mice, and *MCT1* mRNA expression was evaluated by tdTomato fluorescence. In SOD1(G93A) transgenic mice, MCT1 reporter continued to exclusively label oligodendroglia throughout the brain, although there was a marked reduction in MCT1 expression in the ventral horn grey matter (outlined by dashes, Fig. 6d, e) as compared with age-matched control mice (Fig. 6f, g). Reduced MCT1 reporter expression was not due to less oligodendroglia, because CC-1-positive oligodendroglia were preserved, although again CC-1 immunoreactivity may not be labelling fully mature oligodendroglia (Fig. 6e). Taken together, these results suggest that alterations in oligodendrocyte MCT1 may contribute to motoneuron degeneration in ALS, and provide a human disease in which downregulation of MCT1 is associated with neurodegeneration.

Discussion

Our results suggest that oligodendroglia support of axons, through MCT1-based transport of lactate (or pyruvate), is crucial for maintaining axon function and neuron survival and seems to be a fundamental property of oligodendroglia. Disruption of MCT1 could hypothetically cause either reduced lactate export out of, or reduced import into, oligodendroglia because the direction of transport is determined by the relative intra- and extracellular concentrations of lactate and hydrogen ions[6]. Our results suggest primarily disruption of lactate export. First, exogenous lactate completely prevents cell loss in organotypic cultures by compensating for reduced lactate export. Second, oligodendroglia do not degenerate in $MCT1^{+/-}$ mice or when exposed to MCT1 inhibitor and glucose deprivation in organotypic spinal cord cultures. Third, our study and others[10] demonstrate that MCT1 is predominantly localized to the myelin sheath around CNS axons, as expected for a transporter from oligodendroglia to axons. Taken together, our results suggest that MCT1-regulated lactate export from oligodendroglia is a crucial component of the local energy supply to axons, and the disruption of this transport leads to axon dysfunction and ultimately to neuron degeneration. Oligodendrocyte support of axons was also suggested by a recent paper that found increased brain lactate in transgenic mice with a selective deficit in oligodendroglia mitochondrial function[5]. Of course, lactate may also be imported into oligodendroglia and contribute to the production of myelin, as has been shown in dissociated[38] and cortical slice cultures[39], but attenuation of MCT1 does not seem to cause neuron or axon degeneration through this mechanism.

Our results contribute to the emerging data that oligodendroglia are implicated in ALS pathogenesis. Spinal cord ventral grey matter oligodendrocyte precursor cells markedly replicate in ALS mouse models[18], possibly in response to oligodendroglial injury (J.D.R., unpublished data), and cytoplasmic inclusions are found in human ALS oligodendroglia[40]. We found that MCT1 is reduced in affected brain regions from patients with ALS and the ventral horn of the spinal cord in mutant SOD1 transgenic mice. These changes could reflect downregulation of the transporter within intact oligodendroglia, or perhaps death of oligodendroglia and replacement with immature oligodendroglia. Newly generated oligodendroglia, labelled with CC-1, do not contain MCT1 (Fig. 6), and would therefore be unable to supply energy metabolites to axons, leading to axon injury and neuron loss. We propose that this mechanism contributes to motoneuron degeneration in ALS. Future experiments will investigate this specifically by determining whether MCT1 downregulation is specific for ALS and whether MCT1 upregulation or transplantation of oligodendroglia can prolong survival in ALS mouse models.

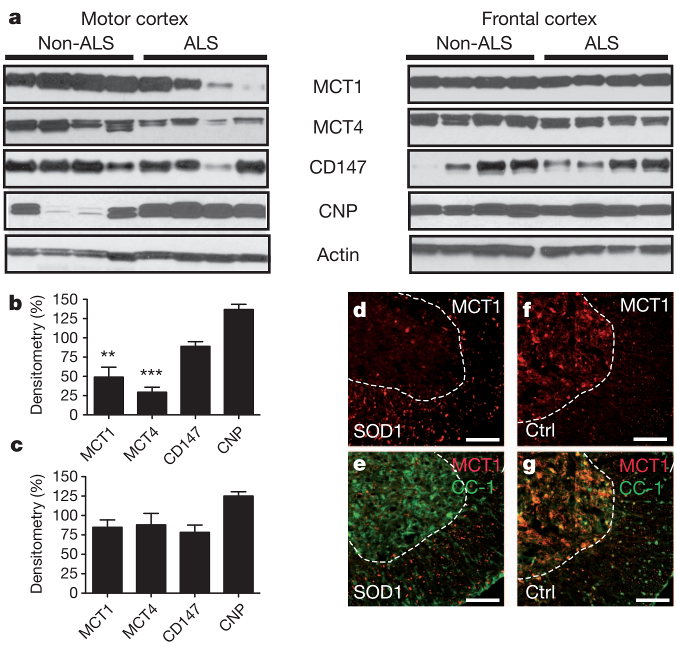


Figure 6 | MCT1 reduced in ALS patients and SOD1(G93A) mice. **a**, Immunoblots of MCT1 and myelin-related proteins from patients with sporadic ALS and non-ALS patients. **b**, **c**, Relative densitometry of proteins from the motor (**b**) and frontal (**c**) cortex ($n = 8$ for each region) of patients with ALS compared with control patients ($n = 9$ for motor and 6 for frontal cortex). Error bars denote s.e.m. **d**–**g**, Immunofluorescence of the MCT1 reporter alone (**d**, **f**; red) and double-labelling with CC-1 (**e**, **g**; green), in end-stage SOD1(G93A) transgenic mice (**d**, **e**) and littermate controls (**f**, **g**). Scale bars, 100 µm. Dashed lines delineate the boundary of ventral horn grey matter.

METHODS SUMMARY

MCT1 BAC mice were developed and MCT1 expression was localized to specific cells by crossing with cell-specific reporter lines, immunostaining for cell-specific markers, or isolating mRNA by FACS and BacTRAP. Critical function of oligodendrocyte MCT1 was evaluated *in vitro* in organotypic spinal cord cultures, and *in vivo* in $MCT1^{+/-}$ or wild-type mice injected with lentiviral vectors. Neuronal toxicity, measured by loss of neurofilament-containing neurons and incorporation of propidium iodide, was provoked in organotypic cultures by treating with ASO or MCT1i. $MCT1^{+/-}$ mice were evaluated by histology, immunohistochemistry and electron microscopy, and compared with littermate controls. For lentiviral experiments, *MCT1* shRNA was subcloned into lentivirus plasmid along with three different promoters (that is, H1, MBP and Cre-dependent V6). Lenti-shRNA, which expresses *MCT1* shRNA driven by the H1 promoter and GFP by the CMV promoter, was injected into the spinal cord of C57Bl6 wild-type mice and motoneurons in the vicinity of virus were counted and compared with control virus injections. Lenti-MBP-shRNA was injected into the optic nerve of Sprague–Dawley rats, and degenerating axons were quantified by electron microscopy and compared with the contralateral optic nerve injected with control virus. Cre-dependent lenti-V6-shRNA was injected into the corpus callosum of PLP1-CreER mice, and axon pathology was assessed by non-phosphorylated neurofilament immunostaining. Finally, MCT1 expression was evaluated by western blots of cortex from patients with ALS and control patients; and MCT1 expression in SOD1(G93A) transgenic mice, obtained from Jackson laboratories, was evaluating by crossing these mice to MCT1 BAC reporter mice.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 March; accepted 13 June 2012.
Published online 11 July 2012.

1. Trapp, B. D. *et al.* Axonal transection in the lesions of multiple sclerosis. *N. Engl. J. Med.* **338,** 278–285 (1998).

2. Garbern, J. Y. *et al.* Patients lacking the major CNS myelin protein, proteolipid protein 1, develop length-dependent axonal degeneration in the absence of demyelination and inflammation. *Brain* **125,** 551–561 (2002).
3. Griffiths, I. *et al.* Axonal swellings and degeneration in mice lacking the major proteolipid of myelin. *Science* **280,** 1610–1613 (1998).
4. Lappe-Siefke, C. *et al.* Disruption of *Cnp1* uncouples oligodendroglial functions in axonal support and myelination. *Nature Genet.* **33,** 366–374 (2003).
5. Fünfschilling, U. *et al.* Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature* **485,** 517–521 (2012).
6. Pierre, K. & Pellerin, L. Monocarboxylate transporters in the central nervous system: distribution, regulation and function. *J. Neurochem.* **94,** 1–14 (2005).
7. Koehler-Stec, E. M., Simpson, I. A., Vannucci, S. J., Landschulz, K. T. & Landschulz, W. H. Monocarboxylate transporter expression in mouse brain. *Am. J. Physiol.* **275,** E516–E524 (1998).
8. Pierre, K., Pellerin, L., Debernardi, R., Riederer, B. M. & Magistretti, P. J. Cell-specific localization of monocarboxylate transporters, MCT1 and MCT2, in the adult mouse brain revealed by double immunohistochemical labeling and confocal microscopy. *Neuroscience* **100,** 617–627 (2000).
9. Halestrap, A. P. & Price, N. T. The proton-linked monocarboxylate transporter (MCT) family: structure, function and regulation. *Biochem. J.* **343,** 281–299 (1999).
10. Rinholm, J. E. *et al.* Regulation of oligodendrocyte development and myelination by glucose and lactate. *J. Neurosci.* **31,** 538–548 (2011).
11. Pellerin, L. & Magistretti, P. J. Glutamate uptake into astrocytes stimulates aerobic glycolysis: a mechanism coupling neuronal activity to glucose utilization. *Proc. Natl Acad. Sci. USA* **91,** 10625–10629 (1994).
12. Walz, W. & Mukerji, S. Lactate production and release in cultured astrocytes. *Neurosci. Lett.* **86,** 296–300 (1988).
13. Pellerin, L. *et al.* Evidence supporting the existence of an activity-dependent astrocyte-neuron lactate shuttle. *Dev. Neurosci.* **20,** 291–299 (1998).
14. Berthet, C. *et al.* Neuroprotective role of lactate after cerebral ischemia. *J. Cereb. Blood Flow Metab.* **29,** 1780–1789 (2009).
15. Suzuki, A. *et al.* Astrocyte-neuron lactate transport is required for long-term memory formation. *Cell* **144,** 810–823 (2011).
16. Benarroch, E. E. Oligodendrocytes: susceptibility to injury and involvement in neurologic disease. *Neurology* **72,** 1779–1785 (2009).
17. Yamanaka, K. *et al.* Astrocytes as determinants of disease progression in inherited amyotrophic lateral sclerosis. *Nature Neurosci.* **11,** 251–253 (2008).
18. Kang, S. H., Fukaya, M., Yang, J. K., Rothstein, J. D. & Bergles, D. E. NG2$^+$ CNS glial progenitors remain committed to the oligodendrocyte lineage in postnatal life and following neurodegeneration. *Neuron* **68,** 668–681 (2010).
19. Pellerin, L., Pellegri, G., Martin, J. L. & Magistretti, P. J. Expression of monocarboxylate transporter mRNAs in mouse brain: support for a distinct role of lactate as an energy substrate for the neonatal vs. adult brain. *Proc. Natl Acad. Sci. USA* **95,** 3990–3995 (1998).
20. Chiry, O. *et al.* Expression of the monocarboxylate transporter MCT1 in the adult human brain cortex. *Brain Res.* **1070,** 65–70 (2006).
21. Gerhart, D. Z., Enerson, B. E., Zhdankina, O. Y., Leino, R. L. & Drewes, L. R. Expression of monocarboxylate transporter MCT1 by brain endothelium and glia in adult and suckling rats. *Am. J. Physiol.* **273,** E207–E213 (1997).
22. Regan, M. R. *et al.* Variations in promoter activity reveal a differential expression and physiology of glutamate transporters by glia in the developing and mature CNS. *J. Neurosci.* **27,** 6607–6619 (2007).
23. Zhou, Q., Wang, S. & Anderson, D. J. Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. *Neuron* **25,** 331–343 (2000).
24. Hanu, R., McKenna, M., O'Neill, A., Resneck, W. G. & Bloch, R. J. Monocarboxylic acid transporters, MCT1 and MCT2, in cortical astrocytes *in vitro* and *in vivo*. *Am. J. Physiol. Cell Physiol.* **278,** C921–C930 (2000).
25. Pellerin, L., Bergersen, L. H., Halestrap, A. P. & Pierre, K. Cellular and subcellular distribution of monocarboxylate transporters in cultured brain cells and in the adult brain. *J. Neurosci. Res.* **79,** 55–64 (2005).
26. Bergersen, L., Rafiki, A. & Ottersen, O. P. Immunogold cytochemistry identifies specialized membrane domains for monocarboxylate transport in the central nervous system. *Neurochem. Res.* **27,** 89–96 (2002).
27. Murray, C. M. *et al.* Monocarboxylate transporter MCT1 is a target for immunosuppression. *Nature Chem. Biol.* **1,** 371–376 (2005).
28. Suh, S. W. *et al.* Astrocyte glycogen sustains neuronal activity during hypoglycemia: studies with the glycogen phosphorylase inhibitor CP-316,819 ([R-R*,S*]-5-chloro-N-[2-hydroxy-3-(methoxymethylamino)-3-oxo-1-(phenylmet hyl)propyl]-1H-indole-2-carboxamide). *J. Pharmacol. Exp. Ther.* **321,** 45–50 (2007).
29. Heyer, E. J., Nowak, L. M. & Macdonald, R. L. Membrane depolarization and prolongation of calcium-dependent action potentials of mouse neurons in cell culture by two convulsants: bicuculline and penicillin. *Brain Res.* **232,** 41–56 (1982).
30. Mayer, M. L. & Westbrook, G. L. Mixed-agonist action of excitatory amino acids on mouse spinal cord neurones under voltage clamp. *J. Physiol. (Lond.)* **354,** 29–53 (1984).
31. Edgar, J. M. *et al.* Early ultrastructural defects of axons and axon-glia junctions in mice lacking expression of Cnp1. *Glia* **57,** 1815–1824 (2009).
32. Morrison, B. M., Shu, I. W., Wilcox, A. L., Gordon, J. W. & Morrison, J. H. Early and selective pathology of light chain neurofilament in the spinal cord and sciatic nerve of G86R mutant superoxide dismutase transgenic mice. *Exp. Neurol.* **165,** 207–220 (2000).
33. Sasaki, S. & Maruyama, S. Increase in diameter of the axonal initial segment is an early change in amyotrophic lateral sclerosis. *J. Neurol. Sci.* **110,** 114–120 (1992).
34. Tekkök, S. B., Brown, A. M., Westenbroek, R., Pellerin, L. & Ransom, B. R. Transfer of glycogen-derived lactate from astrocytes to axons via specific monocarboxylate transporters supports mouse optic nerve activity. *J. Neurosci. Res.* **81,** 644–652 (2005).
35. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101,** 10380–10385 (2004).
36. Doerflinger, N. H., Macklin, W. B. & Popko, B. Inducible site-specific recombination in myelinating cells. *Genesis* **35,** 63–72 (2003).
37. Kirk, P. *et al.* CD147 is tightly associated with lactate transporters MCT1 and MCT4 and facilitates their cell surface expression. *EMBO J.* **19,** 3896–3904 (2000).
38. Sánchez-Abarca, L. I., Tabernero, A. & Medina, J. M. Oligodendrocytes use lactate as a source of energy and as a precursor of lipids. *Glia* **36,** 321–329 (2001).
39. Rinholm, J. E. *et al.* Regulation of oligodendrocyte development and myelination by glucose and lactate. *J. Neurosci.* **31,** 538–548 (2011).
40. Seilhean, D. *et al.* Accumulation of TDP-43 and α-actin in an amyotrophic lateral sclerosis patient with the K17I *ANG* mutation. *Acta Neuropathol.* **118,** 561–573 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Coccia, S. Vidensky, I. Shats, L. Chakravarti, Y. Ayukawa and L. Mamedova for technical support, E. Potter for oligodendrocyte cultures, C. Cooke for electron microscopy, and S. Kang and D. Bergles for providing MOBP–eGFP BAC and CNP BacTrap mice, CASPR and Nav1.6 antibodies. Autopsy specimens were provided by the Johns Hopkins ALS Tissue Bank and the Johns Hopkins University Brain Resource Center supported by National Institutes of Health grants P50AG05146 and PO1NS16375. Support was provided by the Muscular Dystrophy Association (B.M.M. and Yo.L.), NIH-NS33958 (J.D.R.), P2ALS (J.D.R.), Packard Center for ALS (J.D.R.), Human Frontier Science Program-RG118/1998-B (L.P.), Swiss Fonds National de Recherche Scientifique-31003A-125063 (L.P.), Swiss National Science Foundation (FNRS)-3100AO-108336/1 (P.J.M.), Biaggi and Puccini Foundations (P.J.M). We dedicate this manuscript to J. W. Griffin, who passed away while this manuscript was under revision, and are grateful for his contributions to this manuscript and mentorship to many of the authors on this paper.

Author Contributions All of the authors contributed to the design of the experiments. MCT1 BAC reporter experiments were designed and performed by Yo.L., L.J. and P.-W.Z. MCT1 ASO, MCT1i and human western blot experiments were designed and performed by B.M.M., Yu.L., A.T., Yi.L. and J.D.R. Lentiviral experiments were designed and performed by Yo.L., B.M.M., Yu.L. and J.D.R. The heterozygous *MCT1*-null mice were produced by S.L., L.P. and P.J.M., and analysed by Yo.L. Electron miscroscopy work was completed by M.H.F., Yo.L., B.M.M. and J.D.R. Optic nerve lentiviral injections were performed by P.N.H. The manuscript and figures were prepared by B.M.M. and J.D.R. with input from co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.R. (jrothstein@jhmi.edu).

METHODS

Development of MCT1 BAC reporter and MCT1 overexpressor mice. MCT1–tdTomato BAC reporter mice were produced as described previously[22]. The BAC construct was modified to include the entire 19.8 kilobase (kb) *MCT1* gene plus 50 kb upstream of the first exon and 132.2 kb downstream of the last exon (Supplementary Fig. 2a). Founder lines were identified by PCR (primer pair was 5′-CGAGGAGGTCATCAAAGAGT-3′ and 5′-AGAACTTGAGGTGGTC CATC-3′) and also by tdTomato fluorescence within red blood cells. Two expression lines were produced and backcrossed with B6 or C57BL/6 wild-type mice. To produce MCT1 overexpressor mice, the promoter for the astrocyte-specific gene, *Gfap*, was used to drive MCT1 overexpression specifically in astrocytes[41]. *MCT1* was inserted between the *Gfap* promoter sequence and an internal ribosomal entry site (IRES)-eGFP sequence, resulting in expression of MCT1 and eGFP together in astrocytes. Six founder lines were generated, of which five expressed MCT1 and eGFP within CNS astrocytes.

Generation of lines coexpressing the MCT1 BAC reporter and other transgenes. GLT1–eGFP BAC[22], MOBP–eGFP BAC (GENSAT), PLP1–eGFP (Jackson laboratory), PLP1-CreER (Jackson laboratory), CNP–eGFP/RPL10A (Jackson Laboratory), ALDH1L1 (GENSAT), Thy-I–yellow fluorescent protein (YFP) (Jackson laboratory), and SOD1 G93A transgenic mice (Jackson laboratory) were crossed with MCT1 BAC reporter mice to establish double transgenic mice.

FACS, rtPCR and BacTRAP. Brains from 1-month-old MCT1–tdTomato BAC reporter mice were prepared as described previously[42]. Cells were sorted into three groups with a MoFlo MLS high-speed cell sorter (Beckman coulter) running Summit version 4.3 software in the FACS core at The Johns Hopkins University. Total RNA prepared from sorted cells and tissue using Absolutely RNA miniprep kit (Stratagene) was converted to complementary DNA using a cDNA synthesis kit (Applied Biosystems). PCR for *MCT1* or *Actin* (control) was completed using the following primer pairs: *MCT1*, 5′- AAAATGCCACCTGCGATTG GA-3′ and 5′-GCCTGATTAAGTGGAGCCAGG-3′; *Actin,* 5′-AGGCCAACCGTGA AAAGATG-3′ and 5′-CACAGCCTGGATGGCTACGT-3′. TaqMan premade gene-specific probes and 18S ribosomal RNA (as a control) were used for rtPCR. For the FACS isolation of oligodendroglia and astrocytes in the brains, young adult MOBP–eGFP and GLT1–eGFP BAC mice (1–2 months) were used as described previously[42]. For affinity purification of polysomal mRNAs from oligodendroglia and astrocytes in the brain, adult CNP BacTRAP (translating ribosome affinity purification) and ALDH1L1 BacTRAP mice were used, respectively, as described previously[43].

Primary antibodies for western blots and immunohistochemistry. Detailed protocols for western blots and immunohistochemistry are described below. The following primary antibodies were used for western blots: human MCT1 (AbGent; 1:75), mouse MCT1 (Santa Cruz; 1:50), MCT4 (Millipore; 1:400), CD147 (Novus Biologicals; 1:500), CNP (Millipore; 1:1,000), Connexin 43 (Millipore; 1: 500), and β-actin (Millipore; 1:1,000). The following primary antibodies were used for immunofluorescence: mouse MCT1 (Santa Cruz; 1:50), human MCT1 (Santa Cruz; 1:50), OLIG2 (Millipore; 1:500), CC-1 (Calbiochem; 1:50), CNP (Millipore; 1:250), MBP (Covance; 1:250), PDGFRα (BD Pharmingen; 1:100), NG2 (Millipore; 1:100), GFAP (Dako; 1:2,000), IBA1 (Wako; 1:500), NeuN (Millipore; 1:250), SMI-32 (Covance; 1:1,000) ubiquitin (Millipore; 1:250), CASPR (from D. Bergles; 1:1,500), Nav1.6 (from D. Bergles; 1:100), DsRed (Clontech; 1:250), TUJ1 (Millipore; 1:1,000), NeuN (Millipore; 1:1,000) and ALDH1L1 (Neuromab/UC Davis; 1:10).

Lactate uptake assay in oligodendroglioma and astrocytoma cell lines. MO3.13 oligodendroglioma and U87 astrocytoma cell lines are cultured as described in previous reports[44,45]. The lactate uptake assay was completed as described previously[46] with only small modifications. In brief, cells were incubated with 0.5 μCi ml^{-1} L-[1-^{14}C] lactic acid (Perkin-Elmer) in HEPES-buffered, Earl's balanced salt solution (HEBSS) buffer, pH 6.0, containing 150 mM NaCl, 5 mM KCl, 1 mM KH_2PO_4, 0.2 mM $CaCl_2 \cdot 2H_2O$, 3.3 mM MOPS, 10 mM HEPES, 1 mM $MgSO_4 \cdot 7H_2O$. After incubation, uptake was stopped by quickly chilling the cultures to 4 °C. Cells were washed with ice-cold HEPES buffer, homogenized in 0.1 M NaOH and 0.1% Triton X100, and centrifuged at 13,780*g* for 10 min. Radioactivity was measured by scintillation counting and corrected by protein amount.

Immunohistochemistry, histology, light and electron microscopy. Mice were anaesthetized with isoflurane/oxygen and perfused transcardially with 1× PBS followed by 4% paraformaldehyde in PBS (for light microscopy) or 4% paraformaldehyde/2.5% glutaraldehyde in PBS (for electron microscopy), tissue collected, and either cryoprotected in graded concentrations of sucrose and sectioned at 20 μm on a Leica CM1900 cryostat (for light microscopy) or embedded in Epon resin after postfixing in osmium tetroxide (for electron microscopy).

To assess neuropathology by light microscopy, tissue sections were stained with haematoxylin and eosin, eriochrome staining, Bielschowsky silver-staining, or immunohistochemistry. For lectin staining of endothelial cells, biotinylated *Lycopersicon esculentum* lectin (Vector Laboratories; 1:200) was incubated with sections for 1 h. As a secondary antibody, sections were incubated in streptavidin-488 (Vector Laboratories; 1:100). For immunohistochemistry of MCT1, tissue slices were pre-treated with sodium citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6.0) for antigen retrieval (other primary antibodies did not require antigen retrieval) before 1 h incubation at room temperature in blocking buffer (0.3% BSA, 5% skimmed milk and 0.3% Triton-X 100 in PBS). Primary antibodies were incubated overnight at 4 °C, washed and then incubated for 2 h at room temperature in anti-mouse or anti-rabbit IgG Alexa-fluor secondary antibodies (Invitrogen; 1:200). Photomicrographs were taken on Zeiss LSM510 meta confocal microscopy or Zeiss image Z1 fluorescent microscopy. Co-labelled cells with either immunostaining or transgenic reporters were manually counted in Axiovision from digital images taken by fluorescent microscopy. For electron microscopy, 1-mm thick sections were stained for toluidine blue and examined under light microscopy with ×100 oil-immersion objective. Thin sections (70 nm) were obtained and stained for citrate/uranyl acetate. For quantification of axon degeneration in the optic nerves of *MCT1* heterozygous mice and mice treated with lenti-MBP-shRNA, three and ten ×7,000 electron micrographs, respectively, were acquired for each optic nerve using a Zeiss Libra transmission electron microscope. The total number of axons and degenerating axons, as defined by degenerating myelin or dark axons, on each electron micrograph was counted and the percentage of axonal degeneration calculated. All quantification was completed by an investigator blinded to the genotype or treatment group.

Production of cell cultures. Oligodendrocyte cultures were produced from postnatal day (PND) 2–3 mouse cortical tissues of PLP1–eGFP × NG2-tdTomato double transgenic mice. In brief, oligodendrocyte progenitor cells (OPCs) were isolated by FACS and maintained in OPC culture medium (SATO medium containing PDGFRα) for 4 days. OPCs were further differentiated into mature oligodendroglia in the differentiation media (SATO media containing T3 without PDGFRα).

Primary astrocyte cultures were produced from PND 2–3 mouse pups. Cortices were dissected out and dissociated with papain and subsequently cultured on collagen-coated T75 flask in DMEM containing 10% FBS. At 14 days *in vitro* (DIV) astroglial cells were seeded into collagen-coated 6-well plates at a concentration of 7×10^5 cells per well.

Organotypic or cortical spinal cord cultures were produced from PND 7 Sprague–Dawley rat pups or MCT1–tdTomato reporter mice, as described previously[47].

ASOs or MCT1i treatment in organotypic spinal cord cultures. One week after plating organotypic spinal cord cultures, 5 μM MCT1 ASO (IDT) or various doses of MCT1i were added to media for 3 weeks. Sections were either visualized with propidium iodide (7.5 μM, 2 h, Sigma) or fixed with 4% paraformaldehyde, and immunostained with the neuronal marker SMI-32 (Covance; 1:1,000). Large, ventral horn SMI-32-positive neurons were counted by an investigator blinded to the treatment condition.

MCT1i and glucose deprivation in organotypic spinal cord cultures. Two weeks after collection, cultures were pre-treated with MCT1i (ref. 27) or DMSO for 24 h, incubated in propidium iodide for 2 h, and then photographed on a Nikon epifluorescence scope (pre-treatment). Sections were then washed, incubated with glucose-free or normal buffer with or without MCT1i for 2 h, allowed to recover for 2 h, re-incubated with propidium iodide for 2 h, and photographed images were quantified by fluorescent intensity (post-treatment). To examine motoneuron loss, sections were returned to normal media for 2 weeks, fixed in 4% paraformaldehyde, and immunostained with SMI-32 (Covance, 1:1,000). To study cellular localization of propidium iodide, sections were immediately fixed in 4% paraformaldehyde and then processed for immunocytochemistry with primary antibodies. Secondary antibodies used for detection were AlexaFluor 488-conjugated goat anti-rabbit, anti-mouse or anti-chicken antibodies (Invitrogen; 1:1,000).

Organotypic spinal cord cultures treated with bicuculline or glutamate. For bicuculline (BIC) experiment, spinal cord slice cultures (2 weeks post-culture) were treated with 100 μM BIC (or vehicle) in the presence or absence of 1 mM MCT1i for 3 days. To measure cell death, cultures were incubated with propidium iodide and then photographed on a Nikon epifluorescence scope. For glutamate experiment, spinal cord cultures were incubated with MCT1i (or DMSO vehicle for controls) for 24 h, followed by media with or without 250 μM glutamate buffer for 60 min. Propidium iodide staining was performed before and after the glutamate treatment to visualize dead cells. For quantification, fluorescent intensities of pre-treatment sections measured by Image J software for the dorsal

horn in BIC-treated sections, and total spinal cord in glutamate treated sections, were subtracted from post-treatment sections. All values were normalized to untreated control sections.

Production of lentiviral constructs. On the basis of pre-designed ON-TARGETplus SMARTPool siRNA sequences (Dharmacon), DNA oligonucleotides that contain *MCT1* shRNA sequences were synthesized and directly subcloned into pSuper vector (Oligoengine). The sense and antisense oligonucleotide sequences for the most efficient *MCT1* shRNA were as follows: 5′-GATCCC CGTATCATGCTTTACGATTATTCAAGAGATAATCGTAAAGCATGATAC TTTTTTC-3′, 5′-TCGAGAAAAAAGTATCATGCTTTACGATTATCTCTTG AATAATCGTAAAGCATGATACGGG-3′.

DNA fragments spanning the H1 promoter and the shRNA sequences in pSuper construct were amplified by PCR and subcloned into FUGW-CMV to produce lenti-shRNA (in which *MCT1* shRNA is driven by the H1 promoter and GFP by the CMV promoter). Fragments spanning MBP promoter, shRNA and IRES-eGFP sequences were subcloned into FUGW-CMV to produce lenti-MBP-shRNA. V6::*MCT1* shRNA was subcloned into the HpaI and XhoI restriction sites of pSico (Addgene) to produce Cre-inducible lentivirus. Lentiviral constucts, including control GFP constructs, were produced in HEK293 T cells using the FUGW-CMV/Δ8.9/VSVG system.

***MCT1* shRNA lentivirus injected into spinal cord.** Unilateral injections of lenti-shRNA ($n = 10$) or lenti-GFP ($n = 8$) were performed in C57BL6 wild-type mice at 100 days of age at cervical spinal cord levels 4–6 (C4–6), delivering 1.7×10^5 plaque-forming units (p.f.u.) for lenti-GFP or 1.8×10^5 p.f.u. for lenti-shRNA per mouse. In brief, the transverse processes of C4–6 were removed, the dura removed, and three 1-μl injections of lentivirus or media were injected with a 34-gauge Hamilton syringe needle. Animals were euthanized after 4 weeks by transcardial perfusion with 4% paraformaldehyde. Spinal cord sections were cryoprotected, sectioned on a cryostat, and immunostained for neurofilaments (SMI-32), oligodendrocyte-lineage cells (OLIG2), microglia (IBA1) or astrocytes (GFAP), as described earlier. Large (diameter greater than 20 μm) neurofilament-containing neurons in the ventral spinal cord were counted from every third section that localized GFP to the ventral horns on both the virus- and media-injected halves of the spinal cord (mean number of sections counted 11.9 for shRNA group and 12 for GFP group). For each animal, the mean number of motoneurons per section from the virus-injected spinal cord hemisection was divided by the mean number of motoneurons per section from the contralateral media-injected motor neurons per section.

Generation of heterozygous *MCT1*-null mice. Heterozygous *MCT1*-null mice were generated by targeted homologous recombination to replace a 640-base-pair sequence of the *MCT1* gene starting at the translation initiation codon and containing exon 1 as well as part of the first intron with the *LacZ* gene sequence fused with a neomycin (*Neo*) resistance gene sequence and put in frame with the *MCT1* promoter (S. Lengacher *et al.*, manuscript in preparation). Successful recombination event and proper insertion of the *LacZ/Neo* sequence in the targeted locus was controlled by Southern blot on DNA from embryonic stem cells previously electroporated with the targeted vector and selected with gancyclovir and G418. Genotyping of animals was performed by PCR with appropriate set of primers.

***MCT1* shRNA driven by MBP promoter injected into rat optic nerves.** Sixty-day-old Sprague–Dawley rats were anaesthetized with 2% isofluorane, an incision placed in the scalp overlying the orbital ridge, the skin overlying the orbital ridge retracted, and a conjunctival suture placed to provide traction to the globe. The optic nerve was localized, a window cut in the overlying dura, the nerve pierced with a pulled micropipette, and each nerve infused with 3 μl of either lenti-MBP-shRNA or lenti-MBP-GFP (2.4×10^{11} virus particles per ml). After 4 weeks, the rats were transcardially perfused with 4% paraformaldehyde and the optic nerves from the globe to the optic chiasm carefully dissected. The optic nerve near the injection site was post-fixed in 4% paraformaldeyde, cryoprotected in 25% sucrose and cut on a cryostat. These nerve segments were later immunostained with OLIG2, GFAP and SMI-32. Optic nerve distal to the injection was post-fixed in 4% paraformaldehyde, 2.5% glutaraldehyde for 3 days and then processed for electron microscopy as described earlier.

Cre-dependent *MCT1* shRNA lentivirus injected into corpus callosum. Cre-inducible *MCT1* shRNA lentivirus was injected bilaterally into the corpus callosum of PLP-CreER[36], MCT1–tdTomato reporter, and wild-type mice at approximately 60 days of age. Mice were anaesthetized with ketamine and xylazine, placed in a stereotaxic device, and the corpus callosum injected bilaterally with 2 μl lentivirus (2.6×10^{11} virus particles per ml), as previously described[48]. After 1 week, mice were injected intraperitoneally with 1 mg tamoxifen for 5 consecutive days to induce Cre recombination of nucleus-integrated virus. Mice were euthanized 4 weeks later by transcardial perfusion with 4% paraformaldehyde. Sections were cryoprotected, sectioned on a cryostat, and immunostained for neurofilaments (SMI-32) and GFP, as described above.

Western blotting of human autopsy samples, primary cells and mouse tissues. Autopsy samples were obtained from the Johns Hopkins University Brain Resource Center and the Johns Hopkins ALS Tissue Bank. Human or mouse samples were homogenized in TBS with 0.1% SDS, 1% Triton and 10% glycerol, primary cultures were lysed as described previously[49]. Western blotting and densitometry was performed using the enhanced chemilluminescence (ECL) system (GE Healthcare) and NIH Image J.

Statistics. *In-vitro* ASO and MCT1i experiments were analysed using one-way ANOVA followed by Tukey–Kramer post-hoc test. Western blots, rtPCR, and *in-vivo* lenti-shRNA quantifications were analysed using unpaired Student's *t*-tests. Quantification of axon degeneration in lentivirus injections of optic nerve was evaluated with a Mann–Whitney test.

41. Brenner, M., Kisseberth, W. C., Su, Y., Besnard, F. & Messing, A. GFAP promoter directs astrocyte-specific expression in transgenic mice. *J. Neurosci.* **14,** 1030–1037 (1994).
42. Yang, Y. *et al.* Molecular comparison of GLT1$^+$ and ALDH1L1$^+$ astrocytes *in vivo* in astroglial reporter mice. *Glia* **59,** 200–207 (2011).
43. Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135,** 749–762 (2008).
44. Buntinx, M. *et al.* Characterization of three human oligodendroglial cell lines as a model to study oligodendrocyte injury: morphology and oligodendrocyte-specific gene expression. *J. Neurocytol.* **32,** 25–38 (2003).
45. You, F., Osawa, Y., Hayashi, S. & Nakashima, S. Immediate early gene IEX-1 induces astrocytic differentiation of U87-MG human glioma cells. *J. Cell. Biochem.* **100,** 256–265 (2007).
46. Maekawa, F., Minehira, K., Kadomatsu, K. & Pellerin, L. Basal and stimulated lactate fluxes in primary cultures of astrocytes are differentially controlled by distinct proteins. *J. Neurochem.* **107,** 789–798 (2008).
47. Rothstein, J. D., Jin, L., Dykes-Hoberg, M. & Kuncl, R. W. Chronic inhibition of glutamate uptake produces a model of slow neurotoxicity. *Proc. Natl Acad. Sci. USA* **90,** 6591–6595 (1993).
48. McIver, S. R. *et al.* Lentiviral transduction of murine oligodendrocytes *in vivo. J. Neurosci. Res.* **82,** 397–403 (2005).
49. Yang, Y., Gozen, O., Vidensky, S., Robinson, M. B. & Rothstein, J. D. Epigenetic regulation of neuron-dependent induction of astroglial synaptic protein GLT1. *Glia* **58,** 277–286 (2010).

doi:10.1038/nature11255

The 'Higgs' amplitude mode at the two-dimensional superfluid/Mott insulator transition

Manuel Endres[1], Takeshi Fukuhara[1], David Pekker[2], Marc Cheneau[1], Peter Schauβ[1], Christian Gross[1], Eugene Demler[3], Stefan Kuhr[1,4] & Immanuel Bloch[1,5]

Spontaneous symmetry breaking plays a key role in our understanding of nature. In relativistic quantum field theory, a broken continuous symmetry leads to the emergence of two types of fundamental excitation: massless Nambu–Goldstone modes and a massive 'Higgs' amplitude mode. An excitation of Higgs type is of crucial importance in the standard model of elementary particle physics[1], and also appears as a fundamental collective mode in quantum many-body systems[2]. Whether such a mode exists in low-dimensional systems as a resonance-like feature, or whether it becomes overdamped through coupling to Nambu–Goldstone modes, has been a subject of debate[2–9]. Here we experimentally find and study a Higgs mode in a two-dimensional neutral superfluid close to a quantum phase transition to a Mott insulating phase. We unambiguously identify the mode by observing the expected reduction in frequency of the onset of spectral response when approaching the transition point. In this regime, our system is described by an effective relativistic field theory with a two-component quantum field[2,7], which constitutes a minimal model for spontaneous breaking of a continuous symmetry. Additionally, all microscopic parameters of our system are known from first principles and the resolution of our measurement allows us to detect excited states of the many-body system at the level of individual quasiparticles. This allows for an in-depth study of Higgs excitations that also addresses the consequences of the reduced dimensionality and confinement of the system. Our work constitutes a step towards exploring emergent relativistic models with ultracold atomic gases.

Higgs modes are amplitude oscillations of a quantum field and appear as collective excitations in quantum many-body systems as a consequence of spontaneous breaking of a continuous symmetry. Close to a quantum critical point, the low-energy physics of such systems is in many cases captured by an effective Lorentz-invariant critical theory[2]. The minimal version of such a theory describes the dynamics of a complex order parameter $\Psi = |\Psi| e^{i\phi}$ near a quantum phase transition between an ordered ($|\Psi| > 0$) and a disordered phase ($|\Psi| = 0$). Within the ordered phase, the classical energy density has a 'Mexican hat' shape (Fig. 1a) and the order parameter takes on a non-zero value in the minimum of this potential. Its phase, ϕ, thereby acquires a definite value through spontaneous breaking of the rotational symmetry (that is, U(1) symmetry). Expanding the field around the symmetry-broken ground state leads to two types of mode: a Nambu–Goldstone mode and a Higgs mode. These modes are related to phase and amplitude variations of Ψ, respectively (Fig. 1a). In contrast to the phase mode, the amplitude mode has a finite excitation gap (that is, a finite mass), which is expected to show a characteristic softening when approaching the disordered phase (Fig. 1a). The sketched minimal model of an order parameter with $N = 2$ components belongs to a class of O(N) relativistic field theories, which are essential for the study of quantum phase transitions[2].

Despite the fundamental nature of the amplitude mode, a full theoretical understanding of it has not yet been achieved. In particular, the decay of the amplitude mode into lower-lying phase modes, especially in two dimensions, has led to considerable theoretical interest. Specifically, it has been discussed whether a resonance-like feature persists or the decay results in a low-frequency divergence[2–9].

The earliest experimental evidence for a Higgs mode stems from the observation of an unexpected peak in Raman scattering in a superconducting charge density wave compound[10], which was later interpreted as a signal of an amplitude mode[11]. Further examples of experiments in solid-state systems can be found in ref. 6. None of these experiments have studied the mode spectrum across a quantum phase transition, except for neutron scattering experiments on quantum antiferromagnets[12]. In contrast to the work presented here, a resonance-like response of an amplitude mode is expected in these systems, because the phase transition occurs in three dimensions.

Ultracold bosonic atoms in optical lattices offer unique possibilities to study quantum phase transitions in a system with reduced dimensionality[13]. These systems are nearly ideal realizations of the Bose–Hubbard model, which is parameterized by a tunnelling amplitude J and an on-site interaction energy U (Methods). The coupling parameter $j = J/U$ is easily tunable via the lattice depth, and the dimensionality of the system can be reduced by suppressing hopping in a certain direction[14]. At a critical coupling j_c and commensurate filling, the system undergoes a quantum phase transition from a superfluid (ordered) to a Mott insulating (disordered) phase[13], which is described by an O(2) relativistic field theory[2,7]. A number of theoretical works have studied the Higgs mode in this system[7–9,15–19]. In particular, it has been argued that a modulation of the lattice depth can reveal a Higgs mode even in a two-dimensional system[6,8,9].

Previous experiments using a lattice modulation amplitude of 20% were unable to identify the gapped amplitude mode[20,21], most likely owing to the strong drive. A recent theoretical analysis of experiments using Bragg scattering in three-dimensional superfluids interpreted parts of the measured spectrum to be the result of nonlinear coupling to a short-wavelength amplitude mode[22]. Here we experimentally study the long-wavelength and low-energy response, which is described by a relativistic field theory at the quantum critical point.

Our experiment began with the preparation of a two-dimensional, degenerate gas of ^{87}Rb atoms in a single antinode of an optical standing wave[23]. To realize different couplings j, we loaded the two-dimensional gas into a square optical lattice with variable depth V_0 (Fig. 1b). For our trapping parameters and atom numbers (Methods), the density in the centre of the trap is typically one atom per lattice site. We then modulated the lattice depth with an amplitude of 3% at variable frequencies ν_{mod}. The modulation time T_{mod} was set to 20 oscillation cycles ($T_{\mathrm{mod}} = 20/\nu_{\mathrm{mod}}$), thus avoiding an unwanted enhanced response at higher frequencies present in experiments with a fixed modulation time[20,21]. We allowed for an additional hold time such that the sum of modulation and hold time was constant at 200 ms. To quantify the response, we adiabatically increased the lattice depth to reach the atomic limit ($j \approx 0$) and measured the temperature of the

[1]Max-Planck-Institut für Quantenoptik, 85748 Garching, Germany. [2]Department of Physics, California Institute of Technology, Pasadena, California 91125, USA. [3]Physics Department, Harvard University, Cambridge, Massachusetts 02138, USA. [4]University of Strathclyde, SUPA, Glasgow G4 0NG, UK. [5]Ludwig-Maximilians-Universität, 80799 München, Germany.

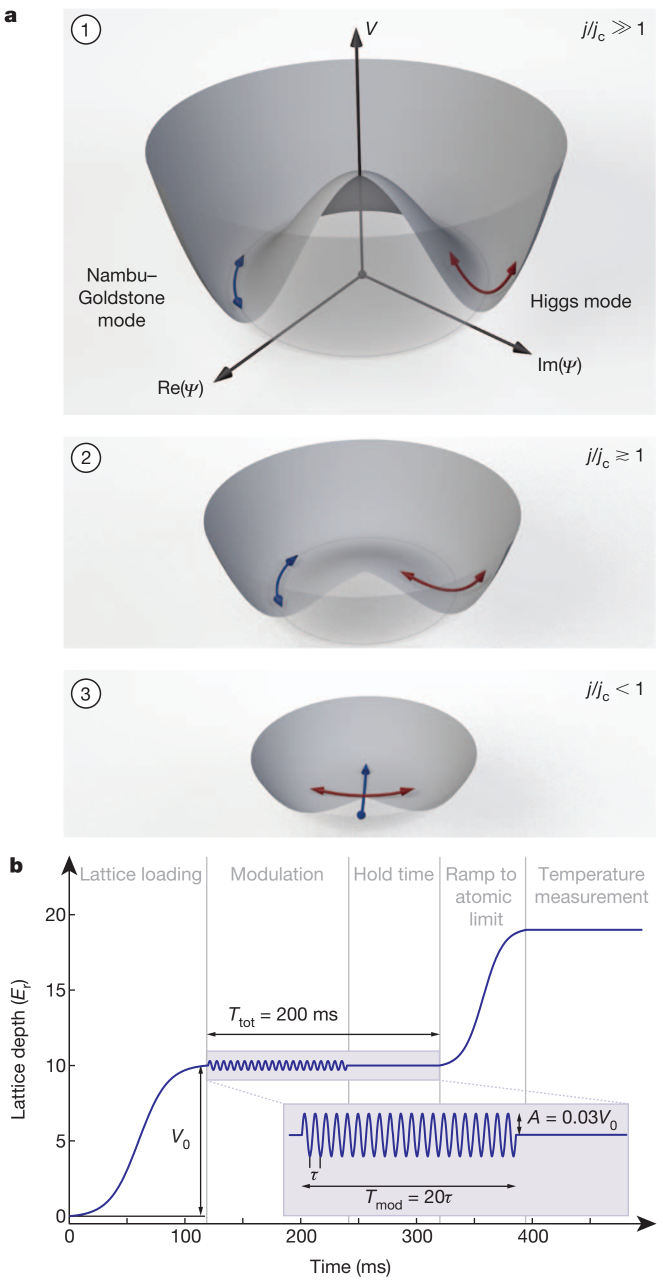


Figure 1 | Illustration of the Higgs mode and experimental sequence. **a**, Classical energy density V as a function of the order parameter Ψ. Within the ordered (superfluid) phase, Nambu–Goldstone and Higgs modes arise from phase and amplitude modulations (blue and red arrows in panel 1). As the coupling $j = J/U$ (see main text) approaches the critical value j_c, the energy density transforms into a function with a minimum at $\Psi = 0$ (panels 2 and 3). Simultaneously, the curvature in the radial direction decreases, leading to a characteristic reduction of the excitation frequency for the Higgs mode. In the disordered (Mott insulating) phase, two gapped modes exist, respectively corresponding to particle and hole excitations in our case (red and blue arrow in panel 3). **b**, The Higgs mode can be excited with a periodic modulation of the coupling j, which amounts to a 'shaking' of the classical energy density potential. In the experimental sequence, this is realized by a modulation of the optical lattice potential (see main text for details). $\tau = 1/\nu_{mod}$; E_r, lattice recoil energy.

system with a recently developed scheme based on single-atom-resolved detection[24]. It is the high sensitivity of this method that allowed us to reduce the modulation amplitude by almost an order of magnitude compared with earlier experiments[20,21] and to stay well within the linear response regime (Supplementary Information).

The results for selected lattice depths V_0 are shown in Fig. 2b. We observe a gapped response with an asymmetric overall shape that will be analysed in the following paragraphs. Notably, the maximum observed temperature after modulation is well below the 'melting' temperature for a Mott insulator in the atomic limit[25], $T_{melt} \approx 0.2U/k_B$ (k_B, Boltzmann's constant), demonstrating that our experiments probe the quantum gas in the degenerate regime. To obtain numerical values for the onset of spectral response, we fitted each spectrum with an error function centred at a frequency ν_0 (Fig. 2b, black lines). With j approaching j_c, the shift of the gap to lower frequencies is already visible in the raw data (Fig. 2b) and becomes even more apparent for the fitted gap ν_0 as a function of j/j_c (Fig. 2a, filled circles). The ν_0 values are in quantitative agreement with a prediction for the Higgs gap ν_{SF} at commensurate filling (solid line):

$$h\nu_{SF}/U = \left[\left(3\sqrt{2}-4\right)\left(1+j/j_c\right)\right]^{1/2}\left(j/j_c-1\right)^{1/2}$$

Here h denotes Planck's constant. This value is based on an analysis of variations around a mean-field state[7,16] (throughout the manuscript, we have rescaled j_c in the theoretical calculations to match the value $j_c \approx 0.06$ obtained from quantum Monte Carlo simulations[26]).

The sharpness of the spectral onset can be quantified by the width of the fitted error function, which is shown as vertical dashed lines in Fig. 2a. Approaching the critical point, the spectral onset becomes sharper, and the width normalized to the centre frequency ν_0 remains constant (Supplementary Fig. 3). The constancy of this ratio indicates that the width of the spectral onset scales with the distance to the critical point in the same way as the gap frequency.

We observe similar gapped responses in the Mott insulating regime (Supplementary Information and Fig. 5a), with the gap closing continuously when approaching the critical point (Fig. 2a, open circles). We interpret this as a result of combined particle and hole excitations with a frequency given by the Mott excitation gap that closes at the transition point[16]. The fitted gaps are consistent with the Mott gap

$$h\nu_{MI}/U = \left[1+\left(12\sqrt{2}-17\right)j/j_c\right]^{1/2}\left(1-j/j_c\right)^{1/2}$$

where ν_{MI} is the Mott gap as predicted by mean-field theory[16] (Fig. 2a, dashed line).

The observed softening of the onset of spectral response in the superfluid regime has led to an identification of the experimental signal with a response from collective excitations of Higgs type. To gain further insight into the full in-trap response, we calculated the eigenspectrum of the system in a Gutzwiller approach[16,22] (Methods and Supplementary Information). The result is a series of discrete eigenfrequencies (Fig. 3a), and the corresponding eigenmodes show in-trap superfluid density distributions, which are reminiscent of the vibrational modes of a drum (Fig. 3b). The frequency of the lowest-lying amplitude-like eigenmode $\nu_{0,G}$ closely follows the long-wavelength prediction for homogeneous commensurate filling ν_{SF} over a wide range of couplings j/j_c until the response rounds off in the vicinity of the critical point due to the finite size of the system (Fig. 3c). Fitting the low-frequency edge of the experimental data can be interpreted as extracting the frequency of this mode, which explains the good quantitative agreement with the prediction for the homogeneous commensurate filling in Fig. 2a. Modes at different frequencies from the lowest-lying amplitude-like mode broaden the spectrum only above the onset of spectral response.

An eigenmode analysis, however, does not yield any information about the finite spectral width of the modes, which stems from the interaction between amplitude and phase excitations. We will consider the question of the spectral width by analysing the low-, intermediate- and high-frequency parts of the response separately. We begin by examining the low-frequency part of the response, which is expected to be governed by a process coupling a virtually excited amplitude mode to a pair of phase modes with opposite momenta. As a result, the response of a strongly interacting, two-dimensional superfluid is

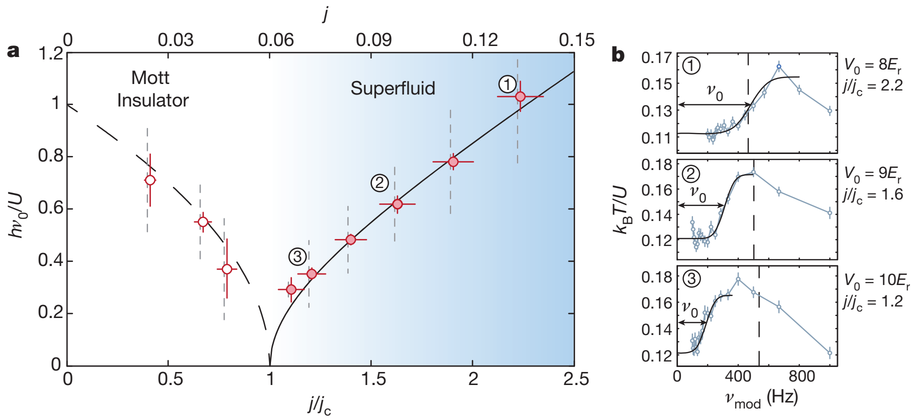


Figure 2 | Softening of the Higgs mode. **a**, The fitted gap values $h\nu_0/U$ (circles) show a characteristic softening close to the critical point in quantitative agreement with analytic predictions for the Higgs and the Mott gap (solid line and dashed line, respectively; see text). Horizontal and vertical error bars denote the experimental uncertainty of the lattice depths and the fit error for the centre frequency of the error function, respectively (Methods). Vertical dashed lines denote the widths of the fitted error function and characterize the sharpness of the spectral onset. The blue shading highlights the superfluid region. **b**, Temperature response to lattice modulation (circles and connecting blue line) and fit with an error function (solid black line) for the three different points labelled in **a**. As the coupling j approaches the critical value j_c, the change in the gap values to lower frequencies is clearly visible (from panel 1 to panel 3). Vertical dashed lines mark the frequency U/h corresponding to the on-site interaction. Each data point results from an average of the temperatures over ~50 experimental runs. Error bars, s.e.m.

expected to diverge at low frequencies, if the probe in use couples longitudinally to the order parameter[2,4,5,9] (for example to the real part of Ψ, if the equilibrium value of Ψ was chosen along the real axis), as is the case for neutron scattering. If, instead, the coupling is rotationally invariant (for example through coupling to $|\Psi|^2$), as expected for lattice modulation, such a divergence could be avoided and the response is expected to scale as ν^3 at low frequencies[3,6,9,17]. Combining this result with the scaling dimensions of the response function for a rotationally symmetric perturbation coupling to $|\Psi|^2$, we expect the low-frequency response to be proportional to $(1 - j/j_c)^{-2}\nu^3$ (ref. 9 and Methods). The experimentally observed signal is consistent with this scaling at the 'base' of the absorption feature (Fig. 4). This indicates that the low-frequency part is dominated by only a few in-trap eigenmodes, which approximately show the generic scaling of the homogeneous system for a response function describing coupling to $|\Psi|^2$.

In the intermediate-frequency regime, it remains a challenge to construct a first-principles analytical treatment of the in-trap system including all relevant decay and coupling processes. Lacking such a theory, we constructed a heuristic model combining the discrete spectrum from the Gutzwiller approach (Fig. 3a) with the line shape for a homogeneous system based on an O(N) field theory in two dimensions, calculated in the large-N limit[3,6] (Methods). An implicit assumption of this approach is a continuum of phase modes, which is

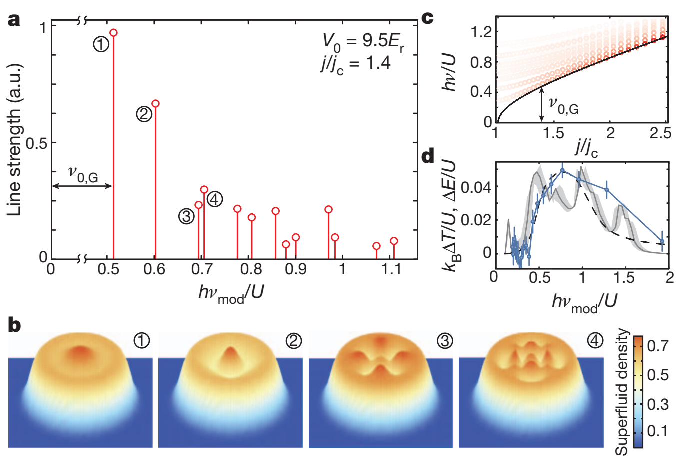


Figure 3 | Theory of in-trap response. **a**, A diagonalization of the trapped system in a Gutzwiller approximation shows a discrete spectrum of amplitude-like eigenmodes. Shown on the vertical axis is the strength of the response to a modulation of j. Eigenmodes of phase type are not shown (Methods) and $\nu_{0,G}$ denotes the gap as calculated in the Gutzwiller approximation. a.u., arbitrary units. **b**, In-trap superfluid density distribution for the four amplitude modes with the lowest frequencies, as labelled in **a**. In contrast to the superfluid density, the total density of the system stays almost constant (not shown). **c**, Discrete amplitude mode spectrum for various couplings j/j_c. Each red circle corresponds to a single eigenmode, with the intensity of the colour being proportional to the line strength. The gap frequency of the lowest-lying mode follows the prediction for commensurate filling (solid line; same as in Fig. 2a) until a rounding off takes place close to the critical point due to the finite size of the system. **d**, Comparison of the experimental response at $V_0 = 9.5E_r$ (blue circles and connecting blue line; error bars, s.e.m.) with a 2 × 2 cluster mean-field simulation (grey line and shaded area) and a heuristic model (dashed line; for details see text and Methods). The simulation was done for $V_0 = 9.5E_r$ (grey line) and for $V_0 = (1 \pm 0.02) \times 9.5E_r$ (shaded grey area), to account for the experimental uncertainty in the lattice depth, and predicts the energy absorption per particle ΔE.

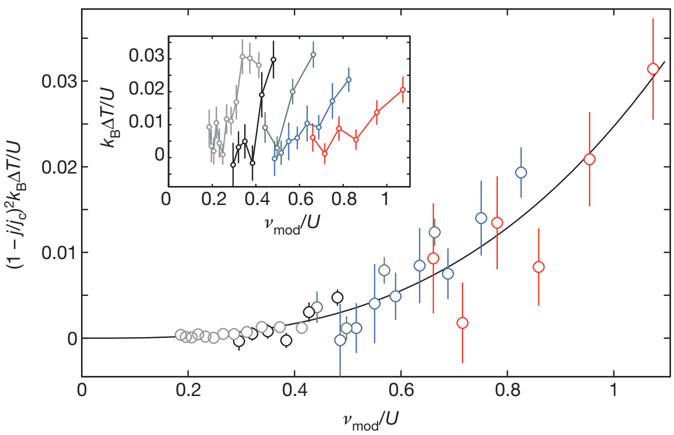


Figure 4 | Scaling of the low-frequency response. The low-frequency response in the superfluid regime shows a scaling compatible with the prediction $(1 - j/j_c)^{-2}\nu^3$ (Methods). Shown is the temperature response rescaled with $(1 - j/j_c)^2$ for $V_0 = 10E_r$ (grey), $9.5E_r$ (black), $9E_r$ (green), $8.5E_r$ (blue) and $8E_r$ (red) as a function of the modulation frequency. The black line is a fit of the form $a\nu^b$ with a fitted exponent $b = 2.9(5)$. The inset shows the same data points without rescaling, for comparison. Error bars, s.e.m.

approximately valid in our case because the frequency spacing between different phase modes is much smaller than the typical gap to the lowest amplitude mode. The model yields quantitative agreement with the low- to intermediate-frequency experimental data for a range of couplings (Fig. 3d, dashed black line, and Supplementary Information), where a relativistic field theoretical treatment of this type is applicable. Furthermore, the response at frequencies more than twice that at the absorption edge remains slightly underestimated.

Part of this high-frequency response might stem from the excitation of several amplitude modes or combinations of amplitude and phase modes, which cannot be described with the Gutzwiller approximation used so far and is only partly captured in the field theoretical treatment. Therefore, we performed a dynamical simulation based on a 2×2 cluster variational wavefunction, which captures the excitation of multiple modes as well as intermode coupling, at least at high momenta. The result is compared with experimental data in Fig. 3d and shows good overall agreement (also compare Fig. 5a with Fig. 5b near the critical point). Notably, the simulation predicts the low-frequency edge, the overall width and the absolute strength of the experimental signal without any fitting parameters. The simulation, which also accounts for fluctuations of the experimental parameters, shows a fine structure that is not observed in the experiment. This indicates that the

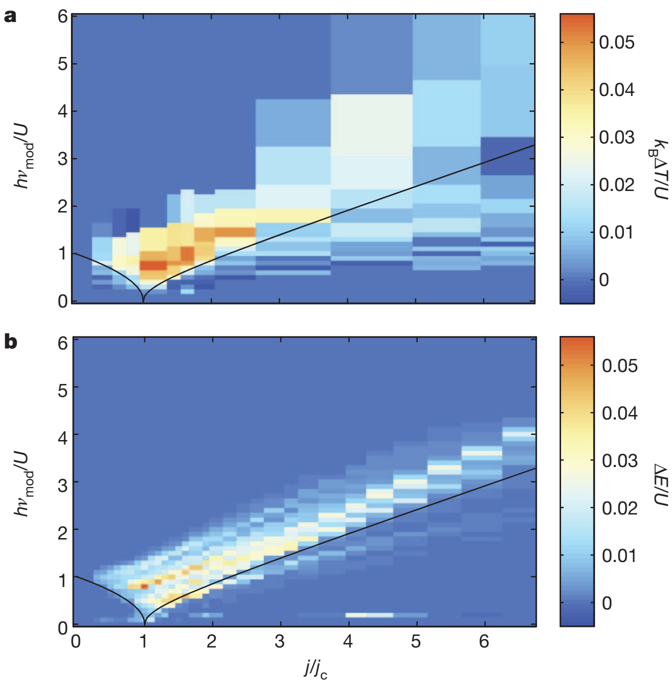


Figure 5 | Response between the strongly interacting limit and the weakly interacting limit. **a**, Change in temperature ΔT as a function of j/j_c and the modulation frequency ν_{mod} in units of U. A pronounced feature close to $j/j_c = 1$ directly shows the existence of the gap and its softening, which is also observed in units of J (Supplementary Fig. 8). As the weakly interacting limit (higher values of j/j_c) is approached, the response broadens and vanishes. **b**, Simulation using a variational 2×2 cluster wavefunction predicting the energy absorption per particle ΔE for the same parameter range. The simulation shows agreement with the experimental data near the critical point in both the softening of the response and the overall width of the absorption band. However, the simulation does not fully reproduce the vanishing of the response at higher j/j_c values. A splitting in the excitation structure at $j/j_c\approx3$ is visible, which might also be present in the experimental data. A low-frequency feature associated with density oscillations at the edges of the trap due to the excitation of phase-like modes is clearly seen in the simulations. This feature occurs below the lowest measured frequency in the experiment and thus is not visible in **a**, (except in the vicinity of the critical point, where the lowest modulation frequencies are close to this feature). Black solid lines show the mean-field predictions as plotted in Fig. 2a.

2×2 cluster treatment still cannot fully capture the broadening of the modes due to coupling to low-energy phase modes.

Our analysis so far has shown the existence of an amplitude mode in the Bose–Hubbard model close to the critical point ($j/j_c\approx1$), where the low-energy description of the system is approximately Lorentz invariant. In the weakly interacting limit ($j/j_c\gg1$), however, the low-energy description (Gross–Pitaevskii theory) becomes effectively Galilean invariant, which precludes the existence of such a mode[27]. To probe the evolution of the amplitude mode response when approaching the weakly interacting limit, we extended our measurements to higher values of the coupling j. The results are shown in Fig. 5a as a density plot, where a pronounced signal for $j/j_c\lesssim3$ directly shows the softening of the mode close to the critical point. As the weakly interacting limit is approached with higher j/j_c values, the response gradually broadens and finally disappears. Despite earlier theoretical treatments of the system in this regime[17,28], a prediction of the disappearance of the response is still lacking. Also, results of the 2×2 cluster variational wavefunction approximation could only partly capture this effect (Fig. 5b).

In conclusion, we have identified and studied long-wavelength Higgs modes in a neutral, two-dimensional superfluid close to the quantum phase transition to a Mott insulating state. This was made possible by recent advances in the high-resolution imaging of single atoms in optical lattices[24,29], leading to a new level of precision for the spectroscopy of ultracold quantum gases. The obtained spectra show softening at the quantum phase transition and are consistent with the generic ν^3 low-frequency scaling for a rotationally invariant coupling to the order parameter in a two-dimensional, strongly interacting superfluid. Furthermore, our results require the development of a quantitative theory valid between the strongly and the weakly interacting regimes capable of predicting the observed disappearance of the response. Our data also call for a first-principles treatment of the discrete nature of Higgs modes in a confined system. In this regard, we note an interesting connection to particle physics, where the spectrum of the Higgs boson in theories with compact extra dimensions[30] may acquire a similar discrete spectrum.

METHODS SUMMARY

The preparation of the two-dimensional, degenerate gas in a single antinode of a vertical optical standing wave is described in ref. 23. The lattice constant for the vertical and both horizontal optical lattices was $a_{lat} = 532$ nm and the trapping frequencies for the two-dimensional system were typically 60 Hz. We calibrated the lattice depths by performing amplitude modulation spectroscopy[23]. The Bose–Hubbard parameters J and U were calculated from the lattice depths by a numerical band-structure calculation[13]. Our systems contained an atom number of 190(36) (where the value in parentheses denotes the standard deviation), resulting in a central density of close to one atom per lattice site. The data point in Fig. 2a at $j/j_c\approx1.1$ was measured with slightly different parameters ($T_{tot} = 300$ ms and $T_{mod} = 15/\nu_{mod}$, where T_{tot} is the sum of modulation time T_{mod} and additional hold time; see Fig. 1b) and is excluded from Figs 4 and 5a. We fitted the temperature response with

$$T = T_0 + \frac{\Delta T}{2}\left\{\mathrm{erf}\left[\frac{1}{\sigma_e}(\nu_{mod}-\nu_0)\right]+1\right\}$$

where erf(x) denotes the error function. The fitting parameters were the temperature offset T_0, the temperature increase ΔT, the width σ_e and the centre frequency ν_0. For the plot in Fig. 4, we chose frequencies in a span from $\nu_0 - 1.5\sigma_e$ to $\nu_0 + 0.5\sigma_e$, with ν_0 and σ_e taken from the error-function fit of the individual responses. To perform the eigenmode analysis in Fig. 3a–c, we first calculated the ground-state wavefunction in a Gutzwiller approximation[16,22] and then linearized the equations of motion. The resulting eigenvalue problem was solved by a Bogoliubov transformation. Furthermore, we separated the Bose–Hubbard Hamiltonian into a time-independent part and a time-dependent part that describes the lattice modulation: $H_{BH} = H_0 + \sin(\omega t)H'$. The line strengths in Fig. 3a, c were calculated using Fermi's golden rule and are proportional to the square of the matrix element of H'. Further details, as well as descriptions of the low-frequency scaling, the heuristic model and the cluster wavefunction approach, can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 April; accepted 16 May 2012.

1. Weinberg, S. *The Quantum Theory of Fields* Vol. 2 (Cambridge Univ. Press, 1996).
2. Sachdev, S. *Quantum Phase Transitions* 2nd edn (Cambridge Univ. Press, 2011).
3. Chubukov, A. V., Sachdev, S. & Ye, J. Theory of two-dimensional quantum Heisenberg antiferromagnets with a nearly critical ground state. *Phys. Rev. B* **49,** 11919–11961 (1994).
4. Sachdev, S. Universal relaxational dynamics near two-dimensional quantum critical points. *Phys. Rev. B* **59,** 14054–14073 (1999).
5. Zwerger, W. Anomalous fluctuations in phases with a broken continuous symmetry. *Phys. Rev. Lett.* **92,** 027203 (2004).
6. Podolsky, D., Auerbach, A. & Arovas, D. P. Visibility of the amplitude (Higgs) mode in condensed matter. *Phys. Rev. B* **84,** 174522 (2011).
7. Altman, E. & Auerbach, A. Oscillating superfluidity of bosons in optical lattices. *Phys. Rev. Lett.* **89,** 250404 (2002).
8. Pollet, L. & Prokof'ev, N. The Higgs mode in a two-dimensional superfluid. Preprint at http://arxiv.org/abs/1204.5190 (2012).
9. Podolsky, D. & Sachdev, S. Spectral functions of the Higgs mode near two-dimensional quantum critical points. Preprint at http://arxiv.org/abs/1205.2700 (2012).
10. Sooryakumar, R. & Klein, M. Raman scattering by superconducting-gap excitations and their coupling to charge-density waves. *Phys. Rev. Lett.* **45,** 660–662 (1980).
11. Littlewood, P. & Varma, C. Gauge-invariant theory of the dynamical interaction of charge density waves and superconductivity. *Phys. Rev. Lett.* **47,** 811–814 (1981).
12. Rüegg, C. *et al.* Quantum magnets under pressure: controlling elementary excitations in TlCuCl3. *Phys. Rev. Lett.* **100,** 205701 (2008).
13. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80,** 885–964 (2008).
14. Spielman, I. B., Phillips, W. D. & Porto, J. V. Mott-insulator transition in a two-dimensional atomic Bose gas. *Phys. Rev. Lett.* **98,** 080404 (2007).
15. Sengupta, K. & Dupuis, N. Mott-insulator-to-superfluid transition in the Bose-Hubbard model: a strong-coupling approach. *Phys. Rev. A* **71,** 033629 (2005).
16. Huber, S. D., Altman, E., Buchler, H. P. & Blatter, G. Dynamical properties of ultracold bosons in an optical lattice. *Phys. Rev. B* **75,** 085106 (2007).
17. Huber, S., Theiler, B., Altman, E. & Blatter, G. Amplitude mode in the quantum phase model. *Phys. Rev. Lett.* **100,** 050404 (2008).
18. Menotti, C. & Trivedi, N. Spectral weight redistribution in strongly correlated bosons in optical lattices. *Phys. Rev. B* **77,** 235120 (2008).
19. Graß, T. D., Santos, F. E. A. & Pelster, A. Real-time Ginzburg-Landau theory for bosons in optical lattices. *Laser Phys.* **21,** 1459–1463 (2011).
20. Stöferle, T., Moritz, H., Schori, C., Köhl, M. & Esslinger, T. Transition from a strongly interacting 1D superfluid to a Mott insulator. *Phys. Rev. Lett.* **92,** 130403 (2004).
21. Schori, C., Stöferle, T., Moritz, H., Köhl, M. & Esslinger, T. Excitations of a superfluid in a three-dimensional optical lattice. *Phys. Rev. Lett.* **93,** 240402 (2004).
22. Bissbort, U. *et al.* Detecting the amplitude mode of strongly interacting lattice bosons by Bragg scattering. *Phys. Rev. Lett.* **106,** 205303 (2011).
23. Endres, M. *et al.* Observation of correlated particle-hole pairs and string order in low-dimensional Mott insulators. *Science* **334,** 200–203 (2011).
24. Sherson, J. F. *et al.* Single-atom-resolved fluorescence imaging of an atomic Mott insulator. *Nature* **467,** 68–72 (2010).
25. Gerbier, F. Boson Mott insulators at finite temperatures. *Phys. Rev. Lett.* **99,** 120405 (2007).
26. Capogrosso-Sansone, B., Söyler, S., Prokof'ev, N. & Svistunov, B. Monte Carlo study of the two-dimensional Bose-Hubbard model. *Phys. Rev. A* **77,** 015602 (2008).
27. Varma, C. Higgs Boson in superconductors. *J. Low Temp. Phys.* **126,** 901–909 (2002).
28. Cazalilla, M. A., Ho, A. F. & Giamarchi, T. Interacting Bose gases in quasi-one-dimensional optical lattices. *N. J. Phys.* **8,** 158 (2006).
29. Bakr, W. S. *et al.* Probing the superfluid-to-Mott insulator transition at the single-atom level. *Science* **329,** 547–550 (2010).
30. Randall, L. & Sundrum, R. An alternative to compactification. *Phys. Rev. Lett.* **83,** 4690–4693 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Weitenberg and J. F. Sherson for their contribution to the design and construction of the apparatus. We thank D. Podolsky, W. Zwerger, S. Sachdev, R. Sensarma, W. Hofstetter, U. Bissbort, L. Pollet and N. Prokof'ev for discussions. We acknowledge funding by the MPG, DFG, EU (NAMEQUAM, AQUTE, Marie Curie Fellowship to M.C.), JSPS (Postdoctoral Fellowship for Research Abroad to T.F.) and California Institute of Technology (IQIM and Lee A. DuBridge fellowship to D.P.). The computations in this paper were done on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University.

Author Contributions All authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.E. (manuel.endres@mpq.mpg.de).

METHODS

Experimental details. The preparation of the two-dimensional, degenerate gas is described in ref. 23. During the experiment, the gas was held in a single antinode of a vertical optical standing wave with a depth of $20(2)E_r$, where E_r denotes the lattice recoil energy $E_r = h^2/8ma_{lat}^2$ with m the atomic mass of ^{87}Rb. The lattice constant for the vertical and both horizontal optical lattices was $a_{lat} = 532$ nm and the trapping frequencies for the two-dimensional system were typically 60 Hz. The ramp for lattice loading and the ramp to the atomic limit were 's' shaped with total durations of 120 and 75 ms, respectively. Our systems had an atom number of 190(36) (where the value in parentheses denotes the standard deviation), resulting in a central density of close to one atom per lattice site. The data point in Fig. 2a at $j/j_c \approx 1.1$ was measured with slightly different parameters ($T_{tot} = 300$ ms and $T_{mod} = 15/\nu_{mod}$, where T_{tot} is the sum of modulation time T_{mod} and additional hold time; see Fig. 1b) and is excluded from Figs 4 and 5a.

Calibration of the lattice depths. We calibrated the lattice depths by performing amplitude modulation spectroscopy[23] with an estimated calibration uncertainty of 1%. To minimize drifts of the lattice depths, we typically repeated this calibration for the horizontal lattice axes after 80 experimental runs. We observed a drift of the lattice depths between these calibrations of maximally 2%, but in most cases no change was observable.

Determination of the Bose–Hubbard parameters. The Bose–Hubbard Hamiltonian is given by

$$H_{BH} = -J\sum_{\langle ik\rangle}\hat{b}_i^\dagger \hat{b}_k + \frac{U}{2}\sum_i \hat{n}_i(\hat{n}_i - 1) + \sum_i (V_i - \mu)\hat{n}_i$$

where $\hat{b}_i^\dagger$ and $\hat{b}_i$ are respectively the boson creation and annihilation operators on lattice site i, $\hat{n}_i$ is the boson number operator, J is the hopping matrix element, U is the on-site interaction, μ is the chemical potential and V_i describes the harmonic trapping potential at site i. The Bose–Hubbard parameters J and U were calculated from the lattice depths by a numerical band-structure calculation[13]. The uncertainties in the coupling $j = J/U$, stated as horizontal error bars in Fig. 2a, result from the experimental uncertainty of the lattice depths of about 2%.

Fit of the mode gap. We fitted the temperature response with

$$T = T_0 + \frac{\Delta T}{2}\left\{\mathrm{erf}\left[\frac{1}{\sigma_e}(\nu_{mod} - \nu_0)\right] + 1\right\}$$

where erf(x) denotes the error function. The fitting parameters were the temperature offset T_0, the temperature increase ΔT, the width σ_e and the centre frequency ν_0. The fit function is a model-free approach to extracting numerical values for the onset of spectral response. The centre frequency ν_0 (circles in Fig. 2a) is a measure of the position of the spectral onset, and the width σ_e (vertical dashed lines in Fig. 2a) is a measure of its sharpness. The width σ_e can also be seen as an estimation of the maximum error in the extraction of the position of the onset. During the least-square optimization, data points at frequencies larger than $\nu_0 + 2.5\sigma_e$ were excluded. The vertical error bar in Fig. 2a is given by the 1σ fitting error for ν_0 and the vertical dashed lines denote $\pm\sigma_e$.

Scaling of the low-frequency response. The amplitude mode response at low frequencies is expected to be proportional to ν^3, which is observed in a weak-coupling expansion[6], in a large N expansion[3,6,9] and in the quantum phase model[17]. Additionally, dimensional analysis shows that the amplitude mode response for an O(N) field theory should follow a scaling of the form[2,6,9] $F(\nu, j/j_c) = A\Delta^{3-2/\nu_c}\Phi(\nu/\Delta)$, where $\Delta \propto (1-j/j_c)^{\nu_c}$ is a typical energy scale, ν_c is the critical exponent associated with Δ, A is a constant and Φ is a universal function. Combining this scaling with the ν^3 prediction yields $F(\nu, j/j_c) \propto A(1-j/j_c)^{-2}\nu^3$ at low frequencies[9]. For the plot in Fig. 4, we chose frequencies in a span from $\nu_0 - 1.5\sigma_e$ to $\nu_0 + 0.5\sigma_e$, with ν_0 and σ_e taken from the error function fit of the individual responses.

Gutzwiller calculation of the eigenmodes in a trap. To perform the eigenmode analysis, we used the Gutzwiller trial wavefunction[16,22]

$$|\psi_{1\times 1}\rangle = e^{i\phi}\prod_i\left(\alpha_i(t)|0\rangle_i + \sqrt{1 - |\alpha_i(t)|^2 - |\gamma_i(t)|^2}|1\rangle_i + \gamma_i(t)|2\rangle_i\right)$$

where $\alpha_i(t)$ and $\gamma_i(t)$ are variational parameters, $|n\rangle_i$ corresponds to a state with n bosons on site i and ϕ is an overall phase. First we obtained the stationary solution $|\psi^0_{1\times 1}\rangle$ (corresponding to $\{\alpha_i^0, \gamma_i^0\}$) by minimizing $\langle\psi^0_{1\times 1}|H_{BH}|\psi^0_{1\times 1}\rangle$ in the entire trap. Next we linearized the equations of motion, which were obtained by minimizing the effective action $\langle\psi_{1\times 1}|i\partial_t - H_{BH}|\psi_{1\times 1}\rangle$ around the stationary solution. The resulting eigenvalue problem was solved by a Bogoliubov transformation $M_{k,ir}$ that relates Bogoliubov creation operators $f_k^\dagger$ (and annihilation operators f_k) to the small fluctuations $\delta\alpha_i$ and $\delta\gamma_i$ around the stationary solution

$$f_k^\dagger = \sum_i \left(M_{k,i1}\delta\alpha_i + M_{i,i2}\delta\alpha_i^* + M_{k,i3}\delta\gamma_i + M_{k,i4}\delta\gamma_i^*\right)$$

where k is the eigenmode index. We can identify the modes as amplitude-like or phase-like using a measure of 'amplitudeness'

$$A = \sum_i M_{k,i1}M_{k,i3}$$

which is positive for amplitude-like modes and negative for phase-like modes (Supplementary Information).

To describe lattice modulation spectroscopy, we separated the Bose–Hubbard Hamiltonian into a time-independent part that describes the system with no modulation and a time-dependent part that describes the lattice modulation: $H_{BH} = H_0 + \sin(\omega t)H'$. The rate of excitation of the kth amplitude or phase mode is given by Fermi's golden rule: $\Gamma(\omega) = \delta(\omega - \omega_k)\left|\langle\psi^0_{1\times 1}f_k|H'|\psi^0_{1\times 1}\rangle\right|^2$. The line strengths plotted in Fig. 3a are proportional to $S_i = \left|\langle\psi^0_{1\times 1}f_k|H'|\psi^0_{1\times 1}\rangle\right|^2$. The line strengths decrease with increasing frequency, because higher-energy modes show short-wavelength spatial variations and do not efficiently couple to lattice modulation.

Heuristic model. We constructed a heuristic model (dashed line in Fig. 3d and Supplementary Information), which combines the frequencies and line strengths of our Gutzwiller calculation with the shape of the response calculated by field theoretical methods. For a given j/j_c value, the Gutzwiller approach yields a series of amplitude-like normal modes with frequencies ν_i and corresponding line strengths S_i. The heuristic model consists of summing up a response function $F(\nu_i, \nu_{mod})$ for each of these frequencies weighted with the corresponding line strengths. A calculation based on a large-N expansion of a two-dimensional O(N) field theory[3,6] yielded a scalar response function for the homogeneous and commensurate system of the form

$$F(\nu, \nu_{mod}) \propto \frac{\nu_{mod}^3}{\left(\nu_{mod}^2 - \nu^2\right)^2 + 4\gamma^2\nu_{mod}^2}$$

A parameterization of the $N = 2$ case of the model can be found in ref. 7 and yields $h\gamma/U = 1/8$. Assuming this response function at each individual normal mode (and measuring all frequencies in units of U/h) results in the final model function

$$F_h(\nu_{mod}) = A_1 + A_2\sum_i S_i\frac{\nu_{mod}^3}{\left(\nu_{mod}^2 - \nu_i^2\right)^2 + 4\gamma^2\nu_{mod}^2}$$

with $\gamma = 1/8$ and fit parameters A_1 and A_2.

Dynamical evolution: 2 × 2 cluster wavefunctions. We performed a study of the dynamical evolution of the system using 2 × 2 cluster variational wavefunctions

$$|\psi_{2\times 2}\rangle = \prod_i\left[a_i(t)\begin{vmatrix}0 & 0\\ 0 & 0\end{matrix}\right\rangle + b_i(t)\begin{vmatrix}1 & 0\\ 0 & 0\end{matrix}\right\rangle + c_i(t)\begin{vmatrix}0 & 1\\ 0 & 0\end{matrix}\right\rangle + \cdots\right]$$

where $a_i(t)$, $b_i(t)$, $c_i(t)$, … are the variational parameters. We restrict the maximum occupation number per site to two. To initialize the dynamics, we obtained the initial trial wavefunction, corresponding to the state of the system before modulation spectroscopy begins, by minimizing $\langle\psi_{2\times 2}|H_{BH}|\psi_{2\times 2}\rangle$. Next we dynamically evolved the trial wavefunction during the modulation drive, the hold time and the ramp to the atomic limit (Fig. 1b). Finally we measured the total energy absorption per particle ΔE of the resulting state (in units of the on-site interaction U in the atomic limit).

LETTER

doi:10.1038/nature11296

Collective bulk carrier delocalization driven by electrostatic surface charge accumulation

M. Nakano[1], K. Shibuya[1]†, D. Okuyama[1], T. Hatano[1], S. Ono[1,2], M. Kawasaki[1,3], Y. Iwasa[1,3] & Y. Tokura[1,3]

In the classic transistor, the number of electric charge carriers—and thus the electrical conductivity—is precisely controlled by external voltage, providing electrical switching capability. This simple but powerful feature is essential for information processing technology, and also provides a platform for fundamental physics research[1–16]. As the number of charges essentially determines the electronic phase of a condensed-matter system, transistor operation enables reversible and isothermal changes in the system's state, as successfully demonstrated in electric-field-induced ferromagnetism[2–4] and superconductivity[5–10]. However, this effect of the electric field is limited to a channel thickness of nanometres or less, owing to the presence of Thomas–Fermi screening. Here we show that this conventional picture does not apply to a class of materials characterized by inherent collective interactions between electrons and the crystal lattice. We prepared metal–insulator–semiconductor field-effect transistors based on vanadium dioxide—a strongly correlated material with a thermally driven, first-order metal–insulator transition well above room temperature[17–23]—and found that electrostatic charging at a surface drives all the previously localized charge carriers in the bulk material into motion, leading to the emergence of a three-dimensional metallic ground state. This non-local switching of the electronic state is achieved by applying a voltage of only about one volt. In a voltage-sweep measurement, the first-order nature of the metal–insulator transition provides a non-volatile memory effect, which is operable at room temperature. Our results demonstrate a conceptually new field-effect device, extending the concept of electric-field control to macroscopic phase control.

Metal–insulator–semiconductor field-effect transistors (MISFETs) work by using the electrostatic charging effect of a capacitor: the number of charges at the topmost surface of a channel material is linearly and reversibly tuned by an external electric field, giving rise to isothermal electrical switching functions. In contrast, a MISFET based on a strongly correlated material (a 'Mott transistor') could have a different operation mechanism[11], in which a small number of electrostatically doped carriers drive all pre-existing localized electrons in the material to be mobile. This occurs by the reduction of the effective Coulomb repulsion energy, as has recently been demonstrated in an FET based on an organic Mott insulator[12]. Several attempts have been made at Mott transistor operation, using correlated oxides such as perovskite manganite[13,14] or nickelate[15,16]; yet none has succeeded in inducing a metallic ground state by an electric field.

In this study, we focus on a classical simple oxide, VO_2. A striking feature of this compound arises from its first-order metal–insulator transition (MIT)[17–20]; above the transition temperature (T_{MI}), the system behaves as a half-filled metal with a $3d^1$ ($S = 1/2$) state, but below T_{MI}, the system adopts the insulating ground state, at which the resistance changes abruptly by several orders of magnitude. The microscopic origin of this transition is the dimerization of V^{4+} ions along the c-axis direction, and resulting lattice transformation from the high-temperature tetragonal (rutile) phase to the low-temperature monoclinic phase, where $3d$ electrons are localized on V sites to form the spin singlet state (Fig. 1a). The driving mechanism of the MIT is still debated, but both strong electron correlation (Mott–Hubbard transition) and electron–lattice coupling (Peierls transition) are generally thought to be important[21–23].

The T_{MI} of VO_2, in bulk (or relaxed thin-film) form, is well above room temperature ($T \sim 340$ K), and can be varied across a wide range by electron-doping[24], or by epitaxial strain in thin films[25]. Figure 1b shows the typical temperature dependence of the resistivity (ρ_{xx}) of relaxed and strained VO_2 films, with a clear thermal hysteresis originating from the first-order nature of the transition. The T_{MI} is high enough to be of practical use; accordingly, many efforts have been devoted to external control of the MIT in VO_2 using, for example, photon irradiation[26,27] or current excitation[28]. The use of a purely electrostatic effect in an FET geometry has also been examined, owing to its great importance for practical applications, but so far, only small changes in the resistance and/or the T_{MI} have been achieved[28–30], possibly because of insufficient electric field available with conventional solid dielectrics.

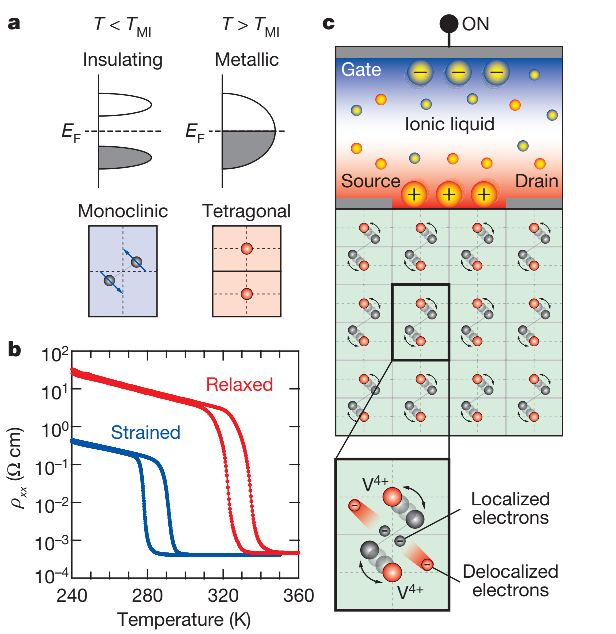


Figure 1 | The first-order metal–insulator transition in VO_2. **a**, Schematic drawings of the thermally driven first-order metal–insulator transition (MIT) in VO_2. **b**, Temperature dependence of the resistivity of strained 10-nm and relaxed 70-nm VO_2 films grown on TiO_2 substrates. **c**, Schematic of an electric-double-layer transistor (EDLT) based on VO_2, potentially enabling electrical switching of the MIT between the metallic tetragonal phase and the insulating monoclinic phase.

[1]Correlated Electron Research Group (CERG) and Cross-correlated Materials Research Group (CMRG), RIKEN Advanced Science Institute, Wako 351-0198, Japan. [2]Central Research Institute of Electric Power Industry, Komae 201-8511, Japan. [3]Quantum-Phase Electronics Center and Department of Applied Physics, University of Tokyo, Tokyo 113-8656, Japan. †Present address: Correlated Electronics Group, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8562, Japan.

To maximize the electron density attainable by applying a static electric field, we have used a recently developed electric-double-layer transistor (EDLT) technique involving an organic ionic liquid, which enables us to tune the surface charge density up to $10^{15}\,cm^{-2}$ (refs 3, 8–10, 15, 16). We fabricated micro-patterned EDLTs with *c*-axis-oriented VO_2 epitaxial thin films[24] in a side-gate configuration (see Methods Summary and Supplementary Information section A for details.) Figure 1c shows a conceptual schematic of the device operation based on a VO_2-EDLT, in which electronic and structural phase transitions are expected to occur simultaneously owing to collective electron–lattice interactions, as demonstrated in femtosecond pump–probe experiments[26,27].

To see the effect of the electric field on the transport properties of VO_2, we first examined this effect on a 10-nm-thick, strained VO_2 film. Figure 2a shows the temperature dependence of the four-terminal normalized sheet resistance (R_s) at different gate voltages (V_G), which were applied in the metallic state above T_{MI}. Below a threshold V_G of ~0.3 V, T_{MI} remains almost unchanged, whereas R_s decreases slightly with increasing V_G at all temperatures, showing normal *n*-type FET behaviour. In this regime, the electric-field effect is limited to the topmost surface of VO_2 by the Thomas–Fermi screening effect, as in conventional MISFETs; hence, T_{MI} cannot be controlled by the electric field because it is governed by the unaffected, bulk part of the film. For $V_G > 0.3$ V, by contrast, T_{MI} decreases dramatically with applied voltage, indicating that the bulk region has been induced to enter the metallic state. At $V_G > 0.7$ V, the MIT vanishes, with the emergence of a metallic ground state. The critical V_G necessary to induce the metallic state is less than 1 V, which is very low by comparison with other EDLT-related studies[3,8–10,15,16]. The shift of T_{MI} is summarized in an electronic phase diagram (Fig. 2a inset). The observed effect is highly reversible and reproducible (see Supplementary Information section B for details). A negative V_G up to -1 V did not affect the transport properties, suggesting that the shift of T_{MI} and the following emergence of the metallic ground state are not driven by the electric field itself (electrostriction effect), but by electrostatic charge accumulation at the surface of the VO_2 channel. Also, the temperature dependence of R_s at $V_G = 0$ V showed an exactly identical curve to that without an ionic liquid, indicating that any strain effect from freezing of the organic ionic liquid used in our experiments is negligible.

The V_G sweep measurement shown in Fig. 2b revealed new behaviours that cannot be achieved in conventional MISFETs. Starting from an insulating (high-*R*) state at $V_G = 0$, R_s decreased nonlinearly by a few orders of magnitude above a threshold V_G of ~1.5 V, clearly showing behaviour characteristic of a Mott transistor[11,12]. In addition, the electric-field-induced metallic (low-*R*) state survived even when V_G was reset to zero, providing a non-volatile memory effect. As the V_G was swept to negative values, R_s increased again, and returned to its initial values at $V_G = 0$, tracing a clear hysteresis loop. Owing to the first-order nature of the MIT, the electronic phase of VO_2 should have a double-minimum potential as a function of the normal coordinate, as schematically illustrated in Fig. 2c. Near T_{MI}, the metallic tetragonal phase and the insulating monoclinic phase are nearly degenerate and hence bistable, giving rise to a thermal hysteresis loop as seen in Fig. 1b. Scanning V_G at fixed temperature corresponds to an isothermal change in potential of one phase with respect to the other, which in principle results in the same effect as the temperature scan. The electric hysteresis loop shown in Fig. 2b thus demonstrates an intrinsic feature of electric-field control of the first-order MIT. The huge resistance ratio (>100:1) between the two persistent states, and the low-voltage-switchable character, suggests excellent potential for practical applications with low power consumption.

Additional signatures of electric-field control of the first-order MIT arise in the thickness dependence of the electric-field effect. Figure 3a shows R_s as a function of temperature for three samples with thicknesses of 10, 20 and 70 nm, for both initial ($V_G = 0$) and electrostatically gated states. Regardless of thickness, all of the films

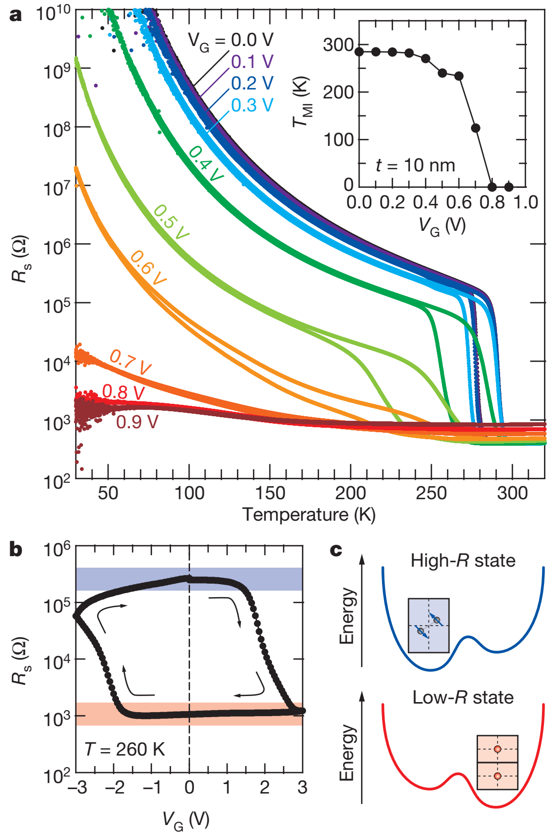


Figure 2 | Effect of electric field on the transport properties of a 10-nm-thick, strained VO_2 film. **a**, Temperature dependence of the sheet resistance (R_s) for a 10-nm strained VO_2 film with different gate voltages (V_G). Inset shows the resulting phase diagram. The transition temperature (T_{MI}) is defined as the average of the two inflection points (for cooling and warming, respectively) in plots of $d[\ln(R_s)]/d(1/T)$ versus temperature[24]. **b**, V_G dependence of R_s, measured at $T = 260$ K. Sweep rate, $15\,mV\,s^{-1}$. **c**, Schematic energy diagrams of VO_2, showing a double-minimum potential as a function of the normal coordinate.

entered the metallic ground state at approximately the same critical V_G, of about 1 V. The sheet conductance (σ_s) of the films in the metallic state increases linearly with increasing thickness, as shown in Fig. 3b, yielding a constant conductivity of the electric-field-induced metallic state of $900\,S\,cm^{-1}$, which is of the same order as that of chemically doped metallic VO_2 films[24]. In both electrostatically and chemically doped materials, the carrier density in the metallic state is confirmed to be $>10^{22}\,cm^{-3}$ by Hall-effect measurements (see Supplementary Information section C for details.) These results show that electrostatic surface charge accumulation can trigger carrier delocalization in a bulk film, irrespective of thickness, beyond the fundamental limitation of the Thomas–Fermi screening effect—distinguishing the present case from the surface-limited electric-field effect seen in conventional MISFETs. The small difference in the critical V_G originates not from the thickness, but from the lattice strain present in the initial ($V_G = 0$) samples; the thinner (10- and 20-nm) films are epitaxially strained, giving rise to a lower T_{MI} than in the relaxed (70-nm) film. The thinner films thus require a smaller V_G to induce the metallic ground state, although the difference is less than a factor of two.

To illustrate the uniqueness of the present situation, we compare the electronic phase diagrams of electrostatically doped 10-, 20- and 70-nm VO_2 films to that of chemically doped 40-nm films[24]. Figure 4 shows the relationship between T_{MI} and the sheet charge density (n_s) of the carriers that are anticipated to be 'additionally' introduced by the respective doping procedures. As mentioned above, the volume carrier density of the metallic states is of the same order ($>10^{22}\,cm^{-3}$) in both

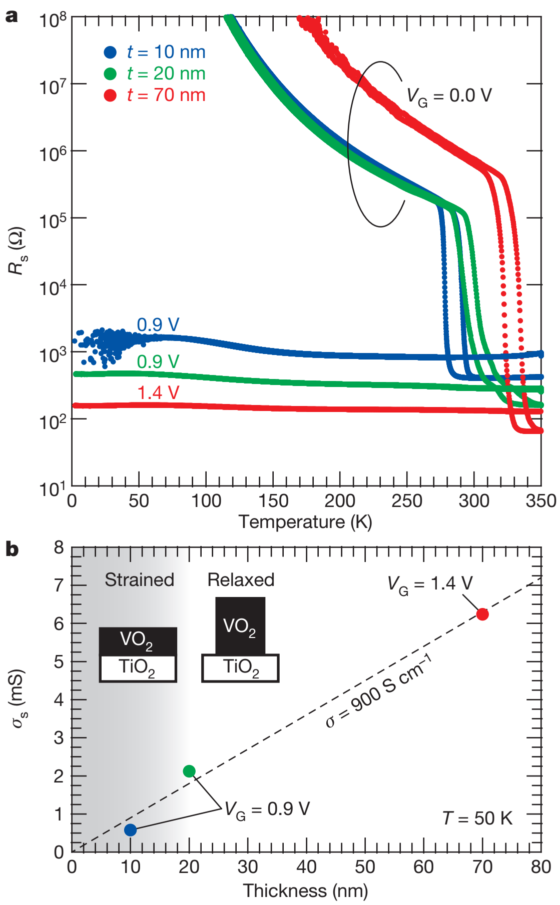


Figure 3 | Emergence of the three-dimensional metallic ground state. **a**, R_s versus temperature in 10-, 20- and 70-nm films, showing both initial ($V_G = 0$) and electric-field-induced metallic states. **b**, Sheet conductance (σ_s) of the electric-field-induced metallic states at $T = 50$ K as a function of film thickness. Inset, schematic depiction of the strain situation in the samples.

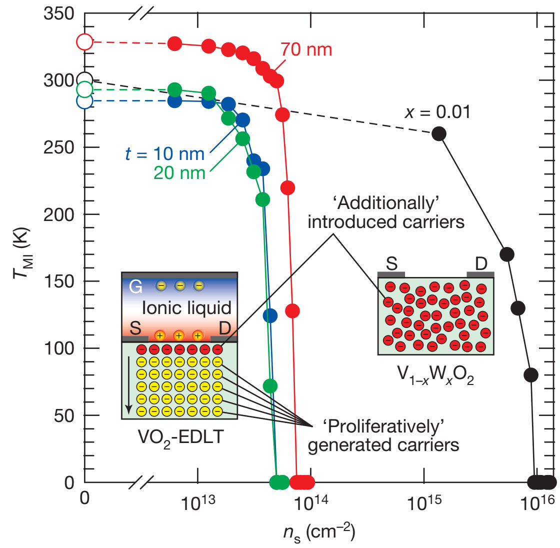


Figure 4 | Electronic phase diagrams of electrostatically and chemically doped VO_2 films. T_{MI} plotted against sheet charge density (n_s) for both VO_2-EDLTs and 40-nm $V_{1-x}W_xO_2$ films[24]. For the electrostatically doped films, n_s was calculated from the capacitor equation, $Q = C_i V_G$ (where Q is the sheet charge and C_i is the areal capacitance), by assuming $C_i = 10\ \mu F\ cm^{-2}$ (ref. 9). For the chemically doped films, n_s was calculated from doping level (x) and film thickness. Insets schematically illustrate the nature of the metallic state in each case, with red and yellow dots representing 'additionally' introduced carriers (that is, n_s) and 'proliferatively' generated carriers, respectively. (See text for details.)

cases. However, as shown in Fig. 4, the values of n_s needed to induce the respective metallic states differ by more than two orders of magnitude. This suggests that, in VO_2-EDLTs, a cascade of phase transitions occurs in a bulk region, suggesting that a simple picture of electrostatic surface charge accumulation, based on a classical capacitor model, is no longer valid. Instead, it seems that surface charge accumulation is accompanied by a collective lattice deformation along the c-axis direction, and resultant delocalization of previously localized electrons in the bulk VO_2 film, leading to a three-dimensional metallic ground state with high carrier density ('proliferatively' generated) throughout the film. In the presence of strong electron correlation and electron–lattice coupling, it is energetically favourable for a system to remain in a single electronic and/or structural phase, at length scales less than a domain size, to minimize interface energy. In fact, a preliminary X-ray diffraction study probing a channel region of a VO_2-EDLT (Supplementary Information section D) did show the occurrence of a structural phase transition on gating, suggesting that the observed bulk phase transition is probably related to this kind of phase separation. Further detailed examinations, including thorough structural analysis, are needed to clarify the underlying microscopic mechanism.

Our results thus provide a new route to controlling the state of matter beyond the fundamental limitation of the screening effect, opening the door to macroscopic phase control by the application of an external voltage. Collective carrier delocalization is one of the most distinctive features of strongly correlated materials, making them good candidates for use in such field-effect device applications. We anticipate that clarifying the mechanism and broadening the scope of relevant materials, based on the concept presented here, will lead to a remarkable discovery of novel electronic phases. For practical applications, we expect that the ability to control macroscopic electronic states by voltage will create useful new device functions—for example, remote transmission of electrical signals over macroscopic length scales or voltage-tunable optical switching.

METHODS SUMMARY

Thin-film growth. VO_2 (001) epitaxial thin films with thickness of 10, 20 and 70 nm were grown on TiO_2 (001) single crystal substrates (Shinkosha Ltd) by pulsed laser deposition at 390 °C under oxygen pressure of 10 mtorr[24]. The films were then annealed at 300 °C for 30 min under 1 torr oxygen to fill oxygen vacancies introduced in the TiO_2 substrate during deposition. From X-ray diffraction measurements, we confirmed that the c-axis-oriented epitaxial films were grown without any impurity phases, and that the 10- and 20-nm films were pseudomorphically strained on the substrates, whereas the 70-nm film was relaxed from epitaxial strain. Thicknesses of the films were evaluated from spacing of X-ray diffraction Laue fringes.

Device fabrication. All of the samples were patterned into a standard Hall-bar geometry, with a side gate fabricated next to a channel by photolithography and argon-ion etching. A device schematic and an optical micrograph of an EDLT are shown in Supplementary Fig. 1a and b, respectively. The dimensions of a channel were 30 μm (or 60 μm in some cases) in width and 520 μm in length. Ti/Au electrodes were deposited by electron-beam evaporation, for both current/voltage probes and a gate electrode. A hard-baked photoresist was used as a separator to electrically isolate gate from channel. After these processes, the samples were annealed at 200 °C in air for several hours to fill oxygen vacancies created in the substrates during device fabrication. Both channel and gate areas were covered just before measurement by a droplet of an organic ionic liquid, N,N-diethyl-N-(2-methoxyethyl)-N-methylammonium bis-trifluoromethylsulphonyl)-imide (DEME-TFSI), with a glass plate on top of the droplet.

Received 28 October 2011; accepted 1 June 2012.

1. Ahn, C. H. *et al.* Electrostatic modification of novel materials. *Rev. Mod. Phys.* **78,** 1185–1212 (2006).
2. Ohno, H. *et al.* Electric-field control of ferromagnetism. *Nature* **408,** 944–946 (2000).
3. Yamada, Y. *et al.* Electrically induced ferromagnetism at room temperature in cobalt-doped titanium dioxide. *Science* **332,** 1065–1067 (2011).

4. Chiba, D. *et al.* Electrical control of the ferromagnetic phase transition in cobalt at room temperature. *Nature Mater.* **10,** 853–856 (2011).
5. Glover, R. E. & Sherrill, M. D. Changes in superconducting critical temperature produced by electrostatic charging. *Phys. Rev. Lett.* **5,** 248–250 (1960).
6. Ahn, C. H. *et al.* Electrostatic modulation of superconductivity in ultrathin $GdBa_2Cu_3O_{7-x}$ films. *Science* **284,** 1152–1155 (1999).
7. Caviglia, A. D. *et al.* Electric field control of the $LaAlO_3/SrTiO_3$ interface ground state. *Nature* **456,** 624–627 (2008).
8. Ueno, K. *et al.* Electric-field-induced superconductivity in an insulator. *Nature Mater.* **7,** 855–858 (2008).
9. Ye, J. T. *et al.* Liquid-gated interface superconductivity on an atomically flat film. *Nature Mater.* **9,** 125–128 (2010).
10. Bollinger, A. T. *et al.* Superconductor-insulator transition in $La_{2-x}Sr_xCuO_4$ at the pair quantum resistance. *Nature* **472,** 458–460 (2011).
11. Newns, D. M. *et al.* Mott transition field effect transistor. *Appl. Phys. Lett.* **73,** 780–782 (1998).
12. Kawasugi, Y. *et al.* Field-induced carrier delocalization in the strain-induced Mott insulating state of an organic superconductor. *Phys. Rev. Lett.* **103,** 116801 (2009).
13. Mathews, S., Ramesh, R., Venkatesan, T. & Benedetto, J. Ferroelectric field effect transistor based on epitaxial perovskite heterostructures. *Science* **276,** 238–240 (1997).
14. Hong, X., Posadas, A., Lin, A. & Ahn, C. H. Ferroelectric-field-induced tuning of magnetism in the colossal magnetoresistive oxide $La_{1-x}Sr_xMnO_3$. *Phys. Rev. B* **68,** 134415 (2003).
15. Asanuma, S. *et al.* Tuning of the metal-insulator transition in electrolyte-gated $NdNiO_3$ thin films. *Appl. Phys. Lett.* **97,** 142110 (2010).
16. Scherwitzl, R. *et al.* Electric-field control of the metal-insulator transition in ultrathin $NdNiO_3$ films. *Adv. Mater.* **22,** 5517–5520 (2010).
17. Morin, F. J. Oxides which show a metal-to-insulator transition at the Neel temperature. *Phys. Rev. Lett.* **3,** 34–36 (1959).
18. Barker, A. S., Verleur, H. W. & Guggenheim, H. J. Infrared optical properties of vanadium dioxide above and below the transition temperature. *Phys. Rev. Lett.* **17,** 1286–1289 (1966).
19. Berglund, C. N. & Guggenheim, H. J. Electronic properties of VO_2 near the semiconductor-metal transition. *Phys. Rev.* **185,** 1022–1033 (1969).
20. Goodenough, J. B. The two components of the crystallographic transition in VO_2. *J. Solid State Chem.* **3,** 490–500 (1971).
21. Wentzcovitch, R. M., Schulz, W. W. & Allen, P. B. VO_2: Peierls or Mott-Hubbard? A view from band theory. *Phys. Rev. Lett.* **72,** 3389–3392 (1994).
22. Rice, T. M., Launois, H. & Pouget, J. P. Comment on "VO_2: Peierls or Mott-Hubbard? A view from band theory". *Phys. Rev. Lett.* **73,** 3042 (1994).
23. Qazilbash, M. M. *et al.* Mott transition in VO_2 revealed by infrared spectroscopy and nano-imaging. *Science* **318,** 1750–1753 (2007).
24. Shibuya, K., Kawasaki, M. & Tokura, Y. Metal-insulator transition in epitaxial $V_{1-x}W_xO_2$ ($0 \leq x \leq 0.33$) thin films. *Appl. Phys. Lett.* **96,** 022102 (2010).
25. Muraoka, Y. & Hiroi, Z. Metal-insulator transition of VO_2 thin films grown on TiO_2 (001) and (110) substrates. *Appl. Phys. Lett.* **80,** 583–585 (2002).
26. Cavalleri, A. *et al.* Femtosecond structural dynamics in VO_2 during an ultrafast solid-solid phase transition. *Phys. Rev. Lett.* **87,** 237401 (2001).
27. Cavalleri, A. *et al.* Band-selective measurements of electron dynamics in VO_2 using femtosecond near-edge X-ray absorption. *Phys. Rev. Lett.* **95,** 067405 (2005).
28. Kim, H. T. *et al.* Mechanism and observation of Mott transition in VO_2-based two- and three-terminal devices. *N. J. Phys.* **6,** 052 (2004).
29. Boriskov, P. P., Velichko, A. A., Pergament, A. L., Stefanovich, G. B. & Stefanovich, D. G. The effect of electric field on metal-insulator phase transition in vanadium dioxide. *Tech. Phys. Lett.* **28,** 406–408 (2002).
30. Ruzmetov, D., Gopalakrishnan, G., Ko, C., Narayanamurti, V. & Ramanathan, S. Three-terminal field effect devices utilizing thin film vanadium oxide as the channel layer. *J. Appl. Phys.* **107,** 114516 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to K. Ueno, H. T. Yuan, J. T. Ye, H. Shimotani, Y. Kasahara and T. Takenobu for discussions, and to T. Arima, H. Ohsumi, S. Takeshita, S. Tardif and Y. Kitagawa for support for X-ray diffraction measurements. This work was supported by the Japan Society for the Promotion of Science (JSAP) through its 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)'. M.N. was supported by RIKEN through the Incentive Research Grant. K.S. was supported by Grants-in-Aid for young scientists (category B, grant number 22760016). Y.I. was partly supported by Grants-in-Aid for Scientific Research (category S, grant number 21224009). The synchrotron X-ray diffraction experiments were performed at BL19LXU in SPring-8 with approval of RIKEN (proposal number 20110091).

Author Contributions M.N. fabricated the devices, performed the measurements and analysed the data. K.S. grew the films. M.N. and D.O. performed the X-ray diffraction measurements under gating. T.H. and S.O. contributed to device fabrications and the experimental setup. M.N., K.S., M.K., Y.I. and Y.T. planned and supervised the study. M.N., M.K., Y.I. and Y.T. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.N. (mnakano@riken.jp) or Y.I. (iwasa@ap.t.u-tokyo.ac.jp).

LETTER

doi:10.1038/nature11215

Structured spheres generated by an in-fibre fluid instability

Joshua J. Kaufman[1], Guangming Tao[1], Soroush Shabahang[1], Esmaeil-Hooman Banaei[1,2], Daosheng S. Deng[3], Xiangdong Liang[4], Steven G. Johnson[4], Yoel Fink[5] & Ayman F. Abouraddy[1]

From drug delivery[1,2] to chemical and biological catalysis[3] and cosmetics[4], the need for efficient fabrication pathways for particles over a wide range of sizes, from a variety of materials, and in many different structures has been well established[5]. Here we harness the inherent scalability of fibre production[6] and an in-fibre Plateau–Rayleigh capillary instability[7] for the fabrication of uniformly sized, structured spherical particles spanning an exceptionally wide range of sizes: from 2 mm down to 20 nm. Thermal processing of a multimaterial fibre[8] controllably induces the instability[9], resulting in a well-ordered, oriented emulsion[10] in three dimensions. The fibre core and cladding correspond to the dispersed and continuous phases, respectively, and are both frozen *in situ* on cooling, after which the particles are released when needed. By arranging a variety of structures and materials in a macroscopic scaled-up model of the fibre, we produce composite, structured, spherical particles, such as core–shell particles, two-compartment 'Janus' particles[11], and multi-sectioned 'beach ball' particles. Moreover, producing fibres with a high density of cores allows for an unprecedented level of parallelization. In principle, 10^8 50-nm cores may be embedded in metres-long, 1-mm-diameter fibre, which can be induced to break up simultaneously throughout its length, into uniformly sized, structured spheres.

Bottom-up approaches to particle synthesis—through nucleation, chemical reactions or self-assembly[12,13]—yield nanometre-scale particles from a variety of materials. The particles produced using such approaches are typically characterized by a large dispersion in the size distribution, and are hampered by coalescence and agglomeration during particle growth. Producing complex multimaterial structures or mono-disperse micro-sized particles is technologically challenging in bottom-up synthesis approaches. On the other hand, top-down approaches[14]—such as microfluidics[15,16], lithography[17,18] and imprint lithography[19]—typically yield larger, mono-disperse particles, but are each suited to a specific material and size range determined by the underlying kinetics of the process. Lithography and imprint lithography produce nanoparticles at low rates, because the produced volume scales with particle size in two-dimensional processes. Furthermore, fabricating structured spherical particles requires non-trivial modifications to these approaches, which ultimately impose constraints on the number of geometric features and types of material combined in a particle. Our new, scalable process exploits a fluid instability occurring within a multimaterial fibre to produce a necklace of spherical particles inside the fibre, with complex geometries combining multiple materials, made with the same ease as producing single-material particles.

Figure 1 outlines our approach to particle generation. The procedure starts with the preparation of a macroscopic scale model, called a 'preform', whose core is assembled from the intended particle constituent materials encased in a supporting cladding. The preform is then thermally drawn into an extended fibre[6,8], until the core diameter approaches the required particle size (Fig. 1a). Figure 1b shows an example of a fibre cross-section, comprising an amorphous, semiconducting chalcogenide glass core (As_2Se_3) encased in a polymer cladding (polyethersulphone, PES); see Supplementary Information section 1 for details. To ensure the integrity of the core, the preform is thermally drawn in a high-viscosity regime ($>10^6$ Pa s), and the fibre emerging from the heating zone is cooled quickly to arrest the development of any axial instability. This drawing process produces axially stable cores[20] with diameters from >1 mm to <3 nm, which sets the upper and lower limits on the potential particle sizes.

Thermal treatment of this fibre induces the Plateau–Rayleigh instability (PRI)[7,9], resulting in the breakup of the continuous core into an orderly arrangement of spherical particles held immobile in the cladding. We capture the full dynamics of the breakup process by applying a temperature gradient along the axis of a fibre section (Fig. 1c). At the low-temperature end (left-hand side), the highly viscous core remains intact. As temperature increases (and viscosity decreases; right-hand side), surface tension dominates and a sinusoidal modulation develops at the core–cladding interface, leading ultimately to the breakup of the core into a string of spherical particles. When we heat the fibre uniformly for sufficient time, the PRI develops globally along its whole length (Supplementary Fig. 1). In contrast to the rapid cooling after fibre drawing, the stationary fibre is maintained at elevated temperature for an extended period of time, allowing the core to break up into uniformly sized droplets which are frozen *in situ* on cooling. The particles may be subsequently released when needed by dissolving the cladding. Our approach is a thermally driven emulsification of dispersed-phase particles (derived from the core) suspended in an immiscible continuous-phase fluid (the cladding)[21]. The core may be considered a pre-filled fluidic channel[15], filled during the construction of the centimetre-scale preform in the solid state, which is reduced to a high-viscosity fluid during the fibre draw and subsequent thermal processing.

This approach to particle fabrication by multimaterial in-fibre emulsification has several important features. First, it can produce uniformly sized spherical particles over an extremely wide range of diameters. This may be appreciated by modelling the fibre core at elevated temperature as a viscous fluid thread surrounded by an infinitely extended viscous immiscible fluid, and making use of the classical Tomotika linear stability theory[22] to calculate the instability growth time τ for potential instability wavelengths λ (see Supplementary Information section 5). We obtain the fastest-growing λ (corresponding to the smallest τ) at which the breakup is favoured to occur for different values of temperature T and core diameter D (Fig. 1d). In general, τ is linearly proportional to D and inversely dependent on T. In this model, a core will break up, at any fixed T, after heating for a sufficiently long time. We have drawn fibres with core diameters ranging from $\sim$2 mm down to 20 nm (Supplementary Information section 1) and used them to produce particles with perfectly spherical and exquisitely smooth-surfaced

[1]CREOL, The College of Optics & Photonics, University of Central Florida, Orlando, Florida 32816, USA. [2]Department of Electrical Engineering & Computer Science, University of Central Florida, Orlando, Florida 32816, USA. [3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

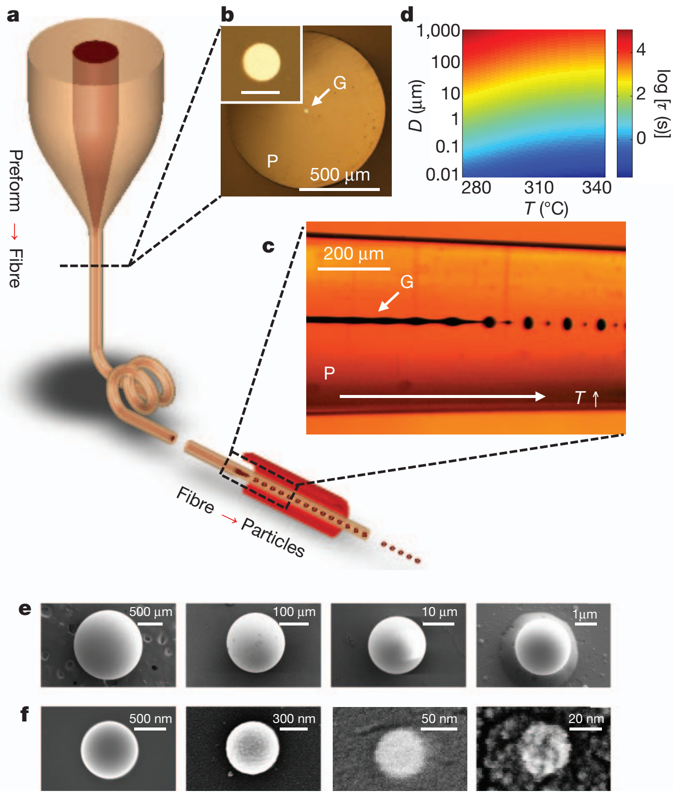


Figure 1 | Fluid capillary instabilities in multimaterial fibres as a route to size-tunable particle fabrication. **a**, A macroscopic preform is thermally drawn into a fibre. Subsequent thermal processing of the fibre induces the PRI, which results in the breakup of the intact core into spherical droplets that are frozen *in situ* on cooling. **b**, Reflection optical micrograph of a fibre cross-section with 20-μm-diameter core; inset shows the core (scale bar, 20 μm). The fibre consists of an As_2Se_3 glass core (G), encased in a PES polymer cladding (P). **c**, Transmission optical micrograph of the fibre side-view in **b** after a temperature (T) gradient is applied along the axis to induce the PRI at the core–cladding interface. **d**, Calculated instability time, τ, for various temperatures T and core diameters D (see Supplementary Information). **e**, SEM images of microparticles with diameters of ~1.4 mm, 200 μm, 18 μm and 2.7 μm. **f**, SEM images of nanoparticles with diameters of ~920, 560, 62 and 20 nm.

external morphology over the entire range of diameters (Fig. 1e, f), confirmed by scanning electron microscope (SEM) imaging after dissolving the polymer cladding using dimethylacetamide[23]. This range corresponds to five orders of magnitude in linear dimension—fifteen in volume, from ~8 mm^3 to ~8,000 nm^3. The polydispersity of the particle distribution for targeted micro- and nanoparticle sizes was determined using dynamic light scattering, and the standard deviation normalized with respect to the mean of the size distribution was found to be ~10% (see Supplementary Fig. 2 for details).

The second key aspect of the in-fibre process is its scalability—that is, the ability to produce large numbers of particles by parallelizing the simultaneous breakup of a high density of cores occupying the same long fibre. Starting from a macroscopic rod, one may in principle convert its entirety into particles of prescribed size. Using a stack-and-draw approach, we have produced fibres of 1 mm outer diameter, containing 12 20-μm cores (Fig. 2a and Supplementary Fig. 3), 4,000 500-nm cores, or 27,000 200-nm cores (Fig. 2d, Supplementary Figs 4, 5; Supplementary Table 1). In principle, one may combine 10^8 50-nm cores in such a fibre with 25% fill factor. This far exceeds the current parallelization capabilities of microfluidics-based approaches[24]. Furthermore, the resulting spatial distribution of particles held immobilized in the scaffold is well-ordered in three dimensions (Fig. 2b, Supplementary Fig. 6). In the axial direction the particles are ordered because the instability growth is dominated by a single wavelength. In the transverse dimensions, order is imposed on the cores during the stacking process (Supplementary Figs 3–6).

The third characteristic is the ease by which this top-down process may be configured to produce structured particles. Because the preform is constructed at the centimetre scale, complex preform geometries may be readily designed and realized, so that the PRI-driven breakup in the drawn fibre produces a desired particle structure. We demonstrate here the size-controllable fabrication of spherical core–shell particles (Fig. 3) and 'Janus' particles (Fig. 4). The preform used to produce the core–shell particles (corresponding to a double emulsion[15]) consists of a polymer core (diameter D_1) and glass cladding (diameter $D_2 \approx 2.5 \times D_1$), surrounded by a polymer matrix (Fig. 3a; cross-sections shown in Fig. 3b, c). The polymer core and glass shell undergo a correlated PRI-driven breakup that results in core–shell particles, observed experimentally (Fig. 3d, e) and confirmed through simulations (Fig. 3f). To confirm that the PRI-driven breakup produces the

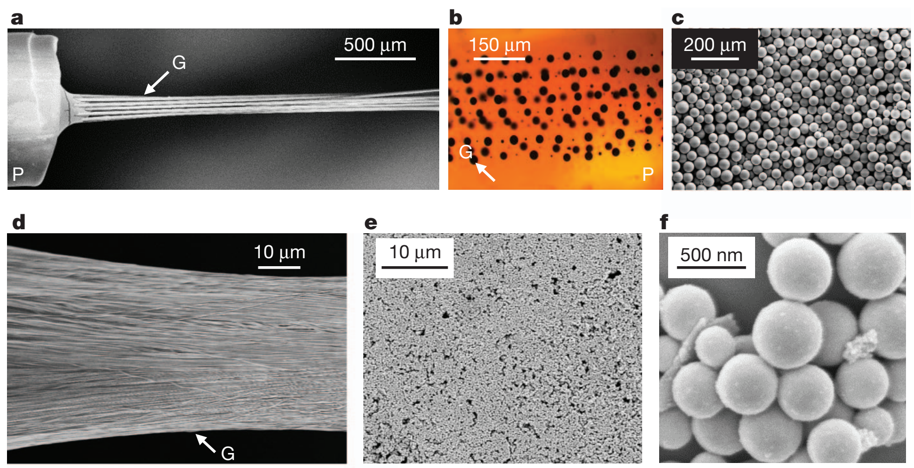


Figure 2 | Scalable fabrication of micro- and nano-scale spherical particles. **a**, SEM micrograph of 12 20-μm intact glass cores (G, As_2Se_3), exposed from a 1-mm-diameter fibre after dissolving the polymer cladding (P, PES). An SEM micrograph of the fibre cross-section is shown in Supplementary Fig. 3. **b**, Transmission optical micrograph of the fibre side-view, showing the cores after global heating of the fibre, which results in the simultaneous breakup of the cores into an ordered distribution of particles in three dimensions held in the polymer cladding. **c**, SEM micrograph of a large number of 40-μm (average diameter) glass particles released from the fibre in **b** by dissolving the polymer cladding. **d**, SEM micrograph of 27,000 200-nm-diameter intact glass cores exposed from a 1-mm-diameter fibre. An SEM micrograph of the fibre cross-section is shown in Supplementary Figs 4, 5. **e**, SEM micrograph of a large number of 400-nm (average diameter) glass particles. **f**, SEM micrograph of a few particles from **e**. See Supplementary Fig. 2 for the particle-size distribution.

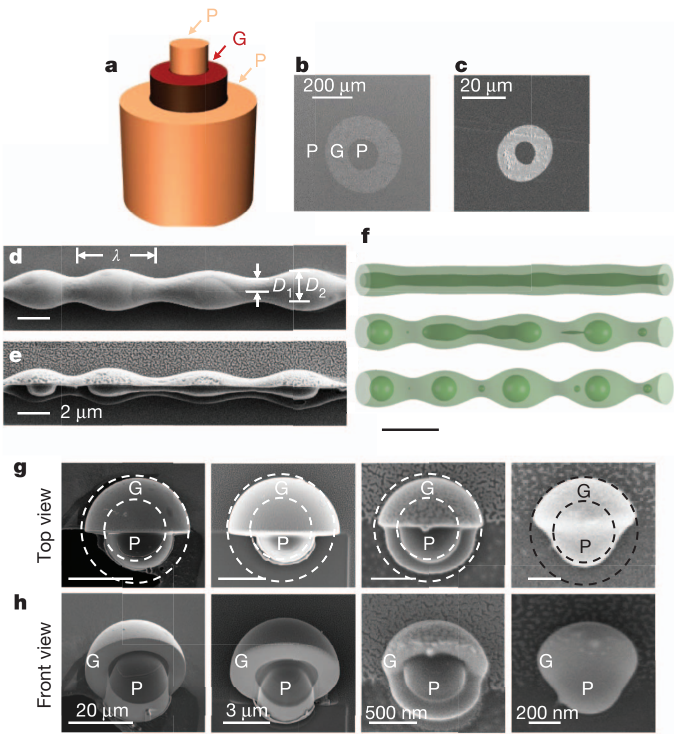


Figure 3 | Polymer-core/glass-shell spherical particle fabrication. **a**, Schematic of the fibre structure (P, G as in Figs 1, 2). **b**, **c**, SEM images of fibre cross-sections. **d**, SEM image of the glass-shell outer surface, showing the modulation characteristic of the PRI. **e**, SEM image of the structure in **d** after sectioning off half of the glass shell using a focused ion beam (FEI 200 THP; current ~10–100 pA), revealing the correlated modulations on the two interfaces (inner polymer/glass and outer glass/polymer interfaces), and resulting ultimately in two concentric spherical surfaces as shown in **g** and **h**. **f**, Three snapshots from a three-dimensional simulation of the Stokes equations using a representative fibre structure (full movie available online; see Supplementary Information), illustrating the full breakup process. Time progresses from top to bottom. Scale bar, 50 μm. Dark green, polymer core; light green, glass shell; the outer polymer scaffold cladding is made transparent for clarity. **g**, Top and **h**, front (tilted) SEM views of four differently sized core–shell particles (outer diameters 34 μm, 7 μm, 1.2 μm and 650 nm, respectively). Scale bars in the corresponding top and front views are the same length.

expected structure, we use a focused ion beam (FIB) to 'slice' the particle down the middle by raster-scanning the FIB across a box with an edge lying through the particle. The FIB etches the semiconducting glass shell, with its higher electrical conductivity, more effectively than the insulating polymer core. Figure 3g, h shows SEM images of particles with outer diameters from 35 μm to 600 nm (the FIB damages the smaller particles that we produced), showing the intact polymer core protruding from the remaining glass half-shell. Figure 3g, h confirms the smooth core/shell interface, the particles' concentric spherical surfaces, and that the expected core/shell diameter ratio is $D_1'/D_2' = (D_1/D_2)^{2/3}$, as dictated by conservation of volume (where D_1' and D_2' are the particle core and shell diameters, respectively). Since $D_1/D_2 = 0.4$ (Fig. 3b, c), we expect $D_1'/D_2' \approx 0.543$, in close agreement with the measured value of ~0.575 (Fig. 3g).

We quantitatively analyse the breakup process in this nested, cylindrical, multi-fluid structure using linear stability analysis[25] to determine the exponential growth rates of small sinusoidal perturbations. The dominant breakup wavelength is plotted as a function of the core/shell viscosity ratio in Supplementary Fig. 7. The predicted and measured breakup length scales are consistent, within the experimental uncertainties in D_1 and viscosity contrast. To study the dynamics of the full breakup process, we performed full three-dimensional simulations of the Stokes equations (valid here because the Reynolds number is low[26]), using a level-set/spectral method[25]. For illustration purposes we used equal core and shell viscosities (10^5 Pa s, corresponding to $T \approx 270$ °C) and an initial diameter $D_1 = 23$ μm. Three snapshots of the simulation, starting from white-noise initial perturbations, are shown in Fig. 3f (full movie available online). The inner interface breaks up first, as predicted from stability analysis and observed experimentally (Fig. 3d, e), and we also occasionally observe small 'satellite' droplets forming among the larger droplets.

The second structured particle we produce is a broken-symmetry, spherical Janus particle, comprising two hemispheres of different optical glasses (Fig. 4). The preform core is constructed of two half cylinders, each of a different semiconducting glass with distinct complex refractive index: G1 (As_2S_3) and G2 ($(As_2Se_3)_{99}Ge_1$) (Fig. 4a–c; Supplementary Information section 7). The induced breakup produces spherical Janus particles held immobilized with the same orientation in the cladding (Fig. 4d). Figure 4e shows a reflection optical micrograph of a single Janus particle removed from the cladding. We confirm the three-dimensional structure of the particle by optically imaging multiple parallel planes cutting through a particle still embedded in the polymer cladding (Fig. 4f), and correlating the optical images with energy-dispersive X-ray diffraction (EDX) spectral images of the particle cross-section (Fig. 4g) that identify arsenic and sulphur. The measurements confirm the three-dimensional, two-compartment structure of the Janus particle.

Modelling Janus-particle formation is difficult because it involves a point where three fluids meet, so that sophisticated level-set techniques are required to describe the interfaces[27]. The physics of such a contact point is not well understood[28], although it is likely to be less relevant in the Stokes regime[29,30]. Nevertheless, energy considerations yield some qualitative predictions. A large glass–glass surface tension, compared to that between glass and polymer, would make it energetically favourable for the Janus particles to pinch in the centre. On the other hand, for negligible glass–glass tension, if the glass–polymer surface tension were very different for the two glasses, energy would be lowered if one glass were to flow to envelop the other. As neither of these scenarios is observed experimentally (Fig. 4d, e), we can conclude that the observed breakup process is consistent with low glass–glass surface tension and similar glass–polymer tensions. These considerations indicate a general strategy for the construction of particles with even more complex geometry. Furthermore, to form two-component particles, the viscosities of the two materials must be matched; we identify pairs of compatible materials by looking for overlapping softening temperatures.

The two particle structures considered above, the core–shell and two-compartment Janus particles, are prototypical structures from which more complex geometries may be constructed. For example, multilayer particles may be produced using a core consisting of nested cylindrical shells of appropriate thicknesses, and additional azimuthal compartments in the particle result from a core appropriately prepared with azimuthal sections. Furthermore, these two prototypical structures may be combined in the same particle. The power of this approach is highlighted in Fig. 4h–k, which shows the fabrication of a 'beach ball' particle, consisting of six equally sized wedge-shaped sections of alternating materials (G1 and G2). The preform consists of a cylindrical core with six equally sized segments, each subtending a 60° polar angle. More complex particle structures may be produced by judiciously structuring the core.

This process uses thermally compatible material systems dominated by viscous forces and surface tension, such as glasses, polymers, metals above their melting temperature, and liquids. An example of breakup in an all-polymer fibre is shown in Supplementary Fig. 10. Moreover, drawing multimaterial fibres with crystalline semiconductor cores[31] (silicon, germanium and III–V binary compounds) and the synthesis of new materials during fibre drawing[32] indicate the possibility of extending our methodology to a wider range of materials. Finally, fibre fabrication technology produces kilometres of fibre in a few hours[6], with a total core mass of 1 kg that is potentially converted entirely into particles, with each metre of fibre containing up to

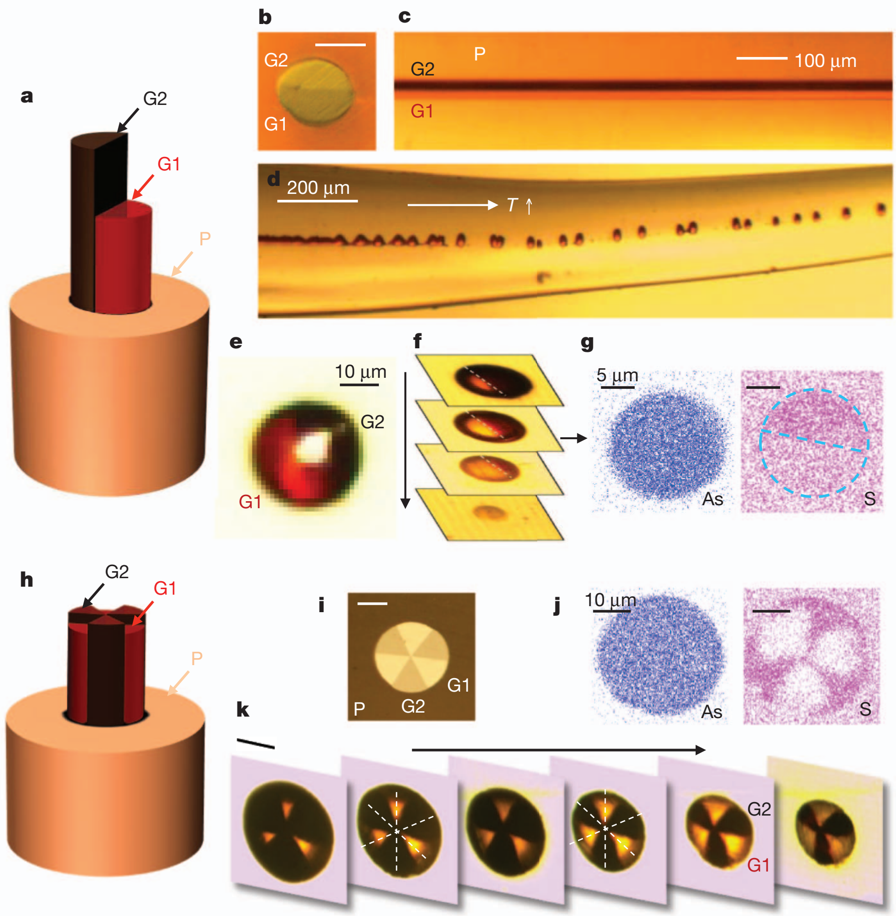


Figure 4 | Broken-symmetry Janus particle and 'beach ball' particle fabrication. **a**, Schematic of the Janus preform. G1, As_2S_3; G2, $(As_2Se_3)_{99}Ge_1$; P, PES. **b**, Reflection optical micrograph of a Janus fibre cross-section; scale bar 20 µm. **c**, Transmission optical micrograph of the fibre side view. **d**, Transmission optical micrograph showing PRI growth, leading to breakup of the Janus particles. **e**, Reflection optical micrograph of an individual Janus particle after removal from the fibre. **f**, Optical micrographs of multiple sections at different depths within a single Janus particle embedded in the fibre, exposed sequentially by polishing. The particle symmetry plane is tilted with respect to the direction of polishing, and the tilt is similar to that in the particle shown in **e**. **g**, EDX spectral images (for arsenic, As, and sulphur, S) of an exposed Janus particle cross-section, corresponding to a section from **f**. The dashed blue circle and line are visual aids. See Supplementary Fig. 8 for full EDX spectrum, Supplementary Fig. 9a–c for another example of EDX spectral imaging, and Supplementary Fig. 9d–f for a demonstration of Janus particle size-control. **h**, Schematic of the preform to produce 'beach ball' particles; G1, G2 and P as above. **i**, Reflection optical micrograph of a 'beach ball' fibre cross-section; scale bar, 20 µm. **j**, EDX spectral images (as in **g**) of an exposed 'beach ball' particle cross-section. **k**, Transmission optical micrographs of the cross-sections of a 40-µm-diameter particle immobilized in the polymer matrix in the fibre; scale bar 20 µm.

10^{14} 100-nm-diameter particles, well ordered in three dimensions. Particles are produced at the same volume rate regardless of the particle size, as it is inherently a three-dimensional process that relies only on the fibre fill factor.

Further control over the preform construction will result in particles with even more complex structures. This scalable process, in which we assemble disparate components that 'fit' together in size and shape macroscopically for the scalable production of size-tunable structured particles, enables a large range of applications. The well ordered, oriented and immobilized three-dimensional particle distribution in a scaffold (Supplementary Fig. 6) could potentially be used as three-dimensional optical and acoustic meta-materials; the surface-tension-driven smooth spherical surface morphology of the particles enables optical-resonance-based sensitive detection of chemical species and pathogens; and three-dimensional structural control over particles impregnated with drugs could help realize sophisticated controlled-release drug delivery systems.

Received 20 December 2012; accepted 2 May 2012.
Published online 18 July 2012.

1. Timko, B. P. *et al.* Advances in drug delivery. *Annu. Rev. Mater. Res.* **41,** 1–20 (2011).
2. Wang, J., Byrne, J. D., Napier, M. E. & DeSimone, J. M. More effective nanomedicines through particle design. *Small* **7,** 1919–1931 (2011).
3. Bell, A. T. The impact of nanoscience on heterogeneous catalysis. *Science* **299,** 1688–1691 (2003).
4. Souto, E. B. & Müller, R. H. Cosmetic features and applications of lipid nanoparticles. *Int. J. Cosmet. Sci.* **30,** 157–165 (2008).
5. Rotello, V. *Nanoparticles*: *Building Blocks for Nanotechnology* (Springer, 2003).
6. Li, T. (ed.) *Optical Fiber Communications* Vol. 1, *Fiber Fabrication* (Academic, 1985).
7. Eggers, J. & Villermaux, E. Physics of liquid jets. *Rep. Prog. Phys.* **71,** 036601 (2008).
8. Abouraddy, A. F. *et al.* Towards multimaterial multifunctional fibres that see, hear, sense and communicate. *Nature Mater.* **6,** 336–347 (2007).
9. Shabahang, S., Kaufman, J. J., Deng, D. S. & Abouraddy, A. F. Observation of the Plateau-Rayleigh capillary instability in multi-material optical fibers. *Appl. Phys. Lett.* **99,** 161909 (2011).
10. Sjöblom, J. *Encyclopedic Handbook of Emulsion Technology* (Marcel Dekker, 2001).
11. Walther, A. & Müller, A. H. E. Janus particles. *Soft Matter* **4,** 663–668 (2008).
12. Cao, G. *Nanostructures and Nanomaterials*: *Synthesis, Properties and Applications* (Imperial College Press, 2004).
13. Vollath, D. *Nanomaterials*: *An Introduction to Synthesis, Properties and Application* (Wiley-VCH, 2008).
14. Merkel, T. J. *et al.* Scalable shape-specific, top-down fabrication methods for the synthesis of engineered colloidal microparticles. *Langmuir* **26,** 13086–13096 (2010).
15. Utada, A. S. *et al.* Monodisperse double emulsions generated from a microcapillary device. *Science* **308,** 537–541 (2005).
16. Dendukuri, D. & Doyle, P. S. The synthesis and assembly of polymeric microparticles using microfluidics. *Adv. Mater.* **21,** 4071–4086 (2009).

17. Dendukuri, D., Pregibon, D. C., Collins, J., Hatton, T. A. & Doyle, P. S. Continuous-flow lithography for high-throughput microparticle synthesis. *Nature Mater.* **5,** 365–369 (2006).
18. Hernandez, C. J. & Mason, T. G. Colloidal alphabet soup: Monodisperse dispersions of shape-designed LithoParticles. *J. Phys. Chem. C* **111,** 4477–4480 (2007).
19. Rolland, J. P. *et al.* Direct fabrication and harvesting of monodisperse, shape specific nano-biomaterials. *J. Am. Chem. Soc.* **127,** 10096–10100 (2005).
20. Kaufman, J. J. *et al.* Thermal drawing of high-density macroscopic arrays of well-ordered sub-5-nm-diameter nanowires. *Nano Lett.* **11,** 4768–4773 (2011).
21. Nie, Z. H. *et al.* Emulsification in a microfluidic flow-focusing device: Effect of the viscosities of the liquids. *Microfluidics and Nanofluidics* **5,** 585–594 (2008).
22. Tomotika, S. On the instability of a cylindrical thread of a viscous liquid surrounded by another viscous fluid. *Proc. R. Soc. Lond. A* **150,** 322–337 (1935).
23. Deng, D. S. *et al.* In-fiber nanoscale semiconductor filament arrays. *Nano Lett.* **8,** 4265–4269 (2008).
24. Nisisako, T. & Torii, T. Microfluidic large-scale integration on a chip for mass production of monodisperse droplets and particles. *Lab Chip* **8,** 287–293 (2008).
25. Liang, X., Deng, D. S., Nave, J.-C. & Johnson, S. G. Linear stability analysis of capillary instabilities for concentric cylindrical shells. *J. Fluid Mech.* **683,** 235–262 (2011).
26. Deng, D. S., Nave, J.-C., Liang, X., Johnson, S. G. & Fink, Y. Exploration of in-fiber nanostructures from capillary instability. *Opt. Express* **19,** 16273–16290 (2011).
27. Smith, K. A., Solis, F. J. & Chopp, D. L. A projection method for motion of triple junctions by levels sets. *Interfaces Free Bound.* **4,** 263–276 (2002).
28. Dussan, V. E. B. On the spreading of liquids on solid surfaces: static and dynamic contact lines. *Annu. Rev. Fluid Mech.* **11,** 371–400 (1979).
29. de Gennes, P. G. Wetting: statics and dynamics. *Rev. Mod. Phys.* **57,** 827–863 (1985).
30. Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic, 1992).
31. Ballato, J. *et al.* Advancements in semiconductor core optical fiber. *Opt. Fiber Technol.* **16,** 399–408 (2010).
32. Orf, N. D. *et al.* Fiber draw synthesis. *Proc. Natl Acad. Sci. USA* **108,** 4743–4747 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Work at UCF was supported by the US National Science Foundation (award number ECCS-1002295), a Ralph E. Powe Junior Faculty Enhancement Award from the Oak Ridge Associated Universities (ORAU), in part by the US Air Force Office of Scientific Research (AFOSR) under contract FA-9550-12-1-0148, and by CREOL, The College of Optics & Photonics. Work at MIT was supported in part by the Materials Research Science and Engineering Program of the US NSF under award number DMR-0819762, and also in part by the US Army Research Office through the Institute for Soldier Nanotechnologies under contract number W911NF-07-D-0004. We thank Sasha Stolyarov, J. Manuel Perez, Sudipta Seal and Kirk Scammon for assistance. We especially thank M. J. Soileau, B. E. A. Saleh, D. N. Christodoulides and M. Z. Bazant for encouragement and support.

Author Contributions J.J.K., Y.F. and A.F.A. developed and directed the project. S.S. first observed the PRI phenomenon, developed the fibre tapering process and the particle extraction approach, and demonstrated the scale invariance of the PRI and particle extraction strategies. G.T. prepared and characterized all the glasses, carried out the preform extrusions, and produced the 'beach ball' fibre. J.J.K. produced the other preforms and fibres, performed PRI breakup and particle extraction experiments, and carried out the SEM, EDX, FIB and optical imaging and characterization. E.-H.B. aided in choice and characterization of materials and in preparation of the polymers. D.S.D., X.L. and S.G.J. carried out the theoretical calculations and performed the simulations. J.J.K., D.S.D., Y.F. and A.F.A. wrote the paper. All authors contributed to the interpretation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.F.A. (raddy@creol.ucf.edu).

doi:10.1038/nature11205

Reconciling the temperature dependence of respiration across timescales and ecosystem types

Gabriel Yvon-Durocher[1,2]*, Jane M. Caffrey[3], Alessandro Cescatti[4], Matteo Dossena[1], Paul del Giorgio[5], Josep M. Gasol[6], José M. Montoya[6], Jukka Pumpanen[7], Peter A. Staehr[8], Mark Trimmer[1], Guy Woodward[1] & Andrew P. Allen[9]*

Ecosystem respiration is the biotic conversion of organic carbon to carbon dioxide by all of the organisms in an ecosystem, including both consumers and primary producers. Respiration exhibits an exponential temperature dependence at the subcellular and individual levels[1], but at the ecosystem level respiration can be modified by many variables[2–4] including community abundance and biomass[5], which vary substantially among ecosystems[6]. Despite its importance for predicting the responses of the biosphere to climate change, it is as yet unknown whether the temperature dependence of ecosystem respiration varies systematically between aquatic and terrestrial environments. Here we use the largest database of respiratory measurements yet compiled to show that the sensitivity of ecosystem respiration to seasonal changes in temperature is remarkably similar for diverse environments encompassing lakes, rivers, estuaries, the open ocean and forested and non-forested terrestrial ecosystems, with an average activation energy similar to that of the respiratory complex[3] (approximately 0.65 electronvolts (eV)). By contrast, annual ecosystem respiration shows a substantially greater temperature dependence across aquatic (approximately 0.65 eV) versus terrestrial ecosystems (approximately 0.32 eV) that span broad geographic gradients in temperature. Using a model[5] derived from metabolic theory[7], these findings can be reconciled by similarities in the biochemical kinetics of metabolism at the subcellular level, and fundamental differences in the importance of other variables besides temperature—such as primary productivity and allochthonous carbon inputs—on the structure of aquatic and terrestrial biota at the community level.

We assessed variability in the temperature dependence of ecosystem respiration within and among a range of aquatic and terrestrial environments using a global compilation of measurements of respiration from nine distinct ecosystem types that represent entire ecosystems or ecosystem components (Supplementary Information 1). We performed two analyses using these data. First, we assessed the sensitivity of ecosystem respiration to seasonal changes in temperature within sites for each ecosystem type, and quantified its variation among sites and across ecosystem types using daily (hereafter short-term) estimates of flux. Second, we determined the temperature sensitivity of respiration at longer timescales by comparing annual (hereafter long-term) fluxes across sites spanning broad geographic gradients in temperature. We reconcile the similarities and differences in the temperature dependence of ecosystem respiration across timescales and ecosystem types using a model[5] derived from metabolic theory[7].

To determine variation among sites in the seasonal temperature dependence of ecosystem respiration for the nine ecosystem types in our compilation, we fit the short-term respiration data to the Boltzmann–Arrhenius function using linear mixed-effects modelling[8] (see Methods Summary, Supplementary Information 2 and 5):

$$\ln R_s(T) = (\bar{E}_R + \varepsilon_E^s)(1/kT_C - 1/kT) + \overline{\ln R(T_C)} + \varepsilon_R^s \qquad (1)$$

In this expression, $\ln R_s(T)$ is the natural logarithm of respiration rate for some arbitrary site s at absolute temperature T (in kelvin (K)), $\bar{E}_R$ is an average among sites for the apparent activation energy, which characterizes the temperature sensitivity of ecosystem respiration, and k is the Boltzmann constant (8.62×10^{-5} eV K^{-1}). We centred the temperature data using a fixed, arbitrary value ($= 288$ K $= 15$ °C) so that $\overline{\ln R(T_C)}$ corresponds to an average among sites for the rate of ecosystem respiration at 15 °C, $R(T_C)$. We would expect $R(T_C)$ to vary among sites due to factors that affect the availability of reduced carbon substrates to support biomass, including net primary production[9,10] and allochthonous carbon inputs[6,11–13], as well as factors that affect the susceptibility of reduced carbon substrates to decomposition by biota, such as C:N:P stoichiometry[6] and water availability[14]. We would also expect $R(T_C)$ to vary seasonally within a site[2,3], resulting in a deviation of the apparent activation energy from $\bar{E}_R$ (Supplementary Information 2), owing to processes that co-vary with temperature, such as litterfall and nutrient turnover in the water column[15]. To account for these factors in our linear mixed-effects models, we treated the slope and intercept as random variables with averages of $\bar{E}_R$ and $\overline{\ln R(T_C)}$, respectively, and site-specific deviations from these averages of ε_E^s and ε_R^s for each site s.

Analyses of the short-term data revealed marked similarities in the seasonal temperature dependence of ecosystem respiration across all nine ecosystem types (Fig. 1). Estimates of the average apparent activation energy, $\bar{E}_R$, were statistically indistinguishable from each other (likelihood ratio test; $\chi^2_8 = 7.36$, $P = 0.50$), with an average of 0.62 eV (Table 1), which corresponds to a Q_{10}—that is, the proportional increase in respiration per 10 °C rise in temperature—of ~2.5 at 15 °C. Consistent with our model, the apparent activation energy varied between sites, as reflected by the significance of the term used to represent ε_E^s in eight of the nine models (Table 1), but this variation was not systematically different among ecosystem types (Supplementary Information 8 and 9). Recent work indicates that this variability partly reflects localized factors—for example, water availability, productivity, allochthonous carbon input—that seasonally co-vary with respiration, $R(T_C)$, and temperature, and can modulate the apparent temperature sensitivity at the site level[2,3]. Our model yields predictions on how the magnitude of this covariation affects the apparent activation energy at a given site, and thus provides a biological interpretation for differences among sites (Supplementary Information 2).

To set our results in a more general theoretical context, we can explore them further by applying a model derived from metabolic theory[5]. Because metabolic theory relates complex ecosystem-level phenomena to the effects of body mass and temperature on individual-level

[1]School of Biological & Chemical Sciences, Queen Mary University of London, London E1 4NS, UK. [2]Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall TR10 9EZ. UK. [3]Center for Environmental Diagnostics and Bioremediation, University of West Florida, 11000 University Parkway, Pensacola, Florida 32514, USA. [4]European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra I-21027, Italy. [5]Département des sciences biologiques, Université du Québec à Montréal, Montréal, Province of Québec, H2X 3X8, Canada. [6]Institute of Marine Sciences (ICM-CSIC), Pg. Marítim de la Barceloneta, 37-49 E-08003 Barcelona, Spain. [7]University of Helsinki Department of Forest Sciences, PO Box 27, FI-00014 University of Helsinki, Finland. [8]Aarhus University, Institute of Bioscience, Frederiksborgvej 399, PO Box 358, 4000 Roskilde, Denmark. [9]Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia.
*These authors contributed equally to this work.

Table 1 | Estimates of the parameters used to characterize the temperature dependence of ecosystem respiration seasonally and annually

Ecosystem type	Short-term $\bar{E}_R$ (95% CI)	r^2	s.d. ε_E^s	s.d. ε_R^s	Cor[ε_E^s,ε_R^s]	Long-term E (95% CI)	r^2
Terrestrial							
Forests	0.62 (0.56–0.68)	0.56	0.17	0.43	0.33	0.42 (0.26–0.60)	0.46
Non-forests	0.70 (0.57–0.80)	0.63	0.20	0.95	0.26	0.00 (−0.30–0.32)	0.00
Soils	0.65 (0.53–0.68)	0.77	0.35	0.84	0.13	0.32 (0.26–0.37)	0.20
Aquatic							
Lake benthic	0.55 (0.44–0.68)	0.74	0.29	0.50	−0.03	0.64 (0.14–1.04)	0.27
Lake pelagic	0.63 (0.55–0.78)	0.73	0.38	1.65	−0.14	1.06 (0.14–2.15)	0.26
Rivers	0.58 (0.44–0.78)	0.81	0.16	0.62	−0.27	1.08 (0.49–1.60)	0.36
Estuarine benthic	0.63 (0.57–0.69)	0.77	0.10	0.53	0.16	0.93 (0.44–1.37)	0.22
Estuarine pelagic	0.59 (0.51–0.69)	0.87	0.29	1.56	−0.22	0.57 (0.22–0.94)	0.25
Oceanic microbial	0.57 (0.46–0.80)	0.72	NA	1.49	NA	NA	NA
All data combined	0.62 (0.60–0.66)	0.86	0.28	1.17	0.01		NA

A likelihood ratio test failed to detect significant differences in $\bar{E}_R$ among ecosystem types when all data were combined into a single mixed-effects analysis (Supplementary Information 5). The standard deviations (s.d.) of site-specific estimates for ε_E^s (equation (1)) were significantly >0 for the eight ecosystems tested ($P<0.05$), indicating differences among sites in the short-term temperature dependence of ecosystem respiration. The standard deviations of site-specific estimates for ε_R^s (equation (1)) were significantly >0 for all ecosystem types ($P<0.0001$), indicating differences among sites in the rate of ecosystem respiration at fixed temperature, $R(T_C)$. Cor[ε_E^s,ε_R^s] refers to the correlation between the site-specific slope and intercept estimates. NA, not applicable.

metabolism, it is particularly useful for understanding how physiology, community structure and abiotic variables combine to influence ecosystem metabolism[5,7]. Ecosystem respiration per unit area for J organisms in an area of size A is equal to the sum $(=(1/A)\sum_{i=1}^{J} r_i)$ of the individual respiration rates[5,16], r_i. Separately performing this summation for autotrophs and heterotrophs yield expressions that relate total respiratory fluxes of the two groups, $R_A(T)$ and $R_H(T)$ ($\mathrm{g\,C\,m^{-2}\,d^{-1}}$), to individual-level normalizations for respiration, r_o^A and r_o^H ($\mathrm{g^{1-\alpha}\,C\,d^{-1}}$), standing stocks of community biomass, M_A and M_H ($\mathrm{g\,C\,m^{-2}}$), biomass-weighted averages for individual body mass, $\langle m_A^{\alpha-1}\rangle$ and $\langle m_H^{\alpha-1}\rangle$ ($\mathrm{g^{\alpha-1}\,C}$), and the size-dependence of individual respiration, characterized by α (Supplementary Information 3)[5]:

$$R_A(T)=r_o^A M_A\langle m_A^{\alpha-1}\rangle e^{-E_R/kT} \quad (2)$$

$$R_H(T)=r_o^H M_H\langle m_H^{\alpha-1}\rangle e^{-E_R/kT} \quad (3)$$

Ecosystem respiration, $R(T)$, is equal to the sum of these fluxes

$$R(T)=R_A(T)+R_H(T)=R(T_C)e^{E_R(1/kT_C-1/kT)} \quad (4)$$

which implies that ecosystem respiration at fixed temperature, $R(T_C)$, is governed partly by the mass-corrected community biomass of autotrophs $M_A\langle m_A^{\alpha-1}\rangle$ and heterotrophs $M_H\langle m_H^{\alpha-1}\rangle$:

$$R(T_C)=\left(r_o^A M_A\langle m_A^{\alpha-1}\rangle+r_o^H M_H\langle m_H^{\alpha-1}\rangle\right)\,e^{-E_R/kT_C} \quad (5)$$

Remarkably, empirical estimates of $\bar{E}_R$ for all nine ecosystem types included in our analyses of the short-term data were statistically indistinguishable from the range of values expected for E_R, that is, 0.6–0.7 eV, based on the biochemical kinetics of the respiratory complex[1,5]. Thus, our findings suggest that the underlying physiological temperature dependence of ecosystem respiration is essentially the same across diverse aquatic and terrestrial environments, despite considerable variability in taxonomic composition and abiotic variables.

We now consider whether ecosystem-level variables, such as gross primary production and allochthonous carbon, modulate the temperature dependence of annual ecosystem respiration across sites that differ in their annual temperature regimes. Metabolic theory yields quantitative predictions for the temperature dependence of annual respiration across ecosystems in which community biomass and metabolism are constrained by the temperature dependence of photosynthesis[5], which is known to be weaker than that of respiration[5,17–19]. The temperature dependence of photosynthesis, although not exponential, can be approximated using an effective activation energy $E_p\approx 0.32$ eV over broad temperature gradients for both aquatic and terrestrial autotrophs[5,19,20], which corresponds to a Q_{10} of ~1.6 at 15 °C. If annual ecosystem respiration, $\tau\langle R(T)\rangle_\tau$, is limited by gross primary production, $\tau\langle P(T)\rangle_\tau$, over the time interval $\tau=1$ yr, then both quantities can be predicted by integrating short-term gross primary production, $P(T)$ ($\mathrm{g\,C\,m^{-2}\,d^{-1}}$), with respect to temperature variation, $T(t)$, through time, t:

$$\tau\langle R(T)\rangle_\tau\approx\tau\theta\langle P(T)\rangle_\tau=\theta\int_\tau P(T(t))dt= \tau\theta p_o e^{-E_P/kT_C}M_A\langle m_A^{\alpha-1}\rangle\left\langle e^{E_P\left(1/kT_C-1/kT\right)}\right\rangle_\tau \quad (6)$$

where p_o is an individual-level normalization for photosynthesis ($\mathrm{g^{1-\alpha}\,C\,d^{-1}}$), $\left\langle e^{E_P\left(1/kT_C-1/kT\right)}\right\rangle_\tau$ is 'Boltzmann-averaged' temperature kinetics over the time interval $=(1/\tau)\int_\tau e^{E_P\left(1/kT_C-1/kT(t)\right)}dt$ (see Supplementary Information 4), and θ is the fraction of gross primary production respired by autotrophs and heterotrophs. The remaining non-respired fraction, $1-\theta$, may accumulate *in situ*, may be released through alternative pathways such as fire or volatile organic carbon emissions[21], or may be exported as allochthonous carbon to other ecosystems, where it can potentially enhance respiration and heterotrophic biomass production[11,22]. Equation (6) implies a proportional relationship between ecosystem respiration and gross primary production, consistent with broad-scale comparative analyses of terrestrial ecosystems[10,23], based on the assumptions that allochthonous carbon inputs are much smaller than autochthonous primary production[24], and that heterotrophic respiration consumes most net primary production[5,10]. Although both assumptions seem to be reasonable first approximations in terrestrial ecosystems over sufficiently long timescales, for example, centuries[25], imbalances between terrestrial plant production and ecosystem respiration are well established over timescales of years to decades[26]. Nevertheless, equation (6) provides a useful framework for assessing the directions and magnitudes of such imbalances.

Equation (6) predicts that long-term ecosystem respiration should be governed by the temperature dependence of photosynthesis (≈ 0.32 eV) across terrestrial ecosystems, where long-term respiration is limited by gross primary production at the site level[23]. This assumption is clearly violated in many aquatic ecosystems because allochthonous carbon inputs can contribute to heterotrophic biomass production and ecosystem respiration, beyond that which could be supported by autochthonous primary production alone[11,22]. In addition, the fraction of autochthonous primary production consumed locally by heterotrophic respiration generally declines with increasing productivity in aquatic ecosystems[15]. Both of these processes may effectively decouple respiration from photosynthesis at the site level. This leads us to propose that there are fundamental differences in the long-term temperature dependence of ecosystem respiration between aquatic and terrestrial environments.

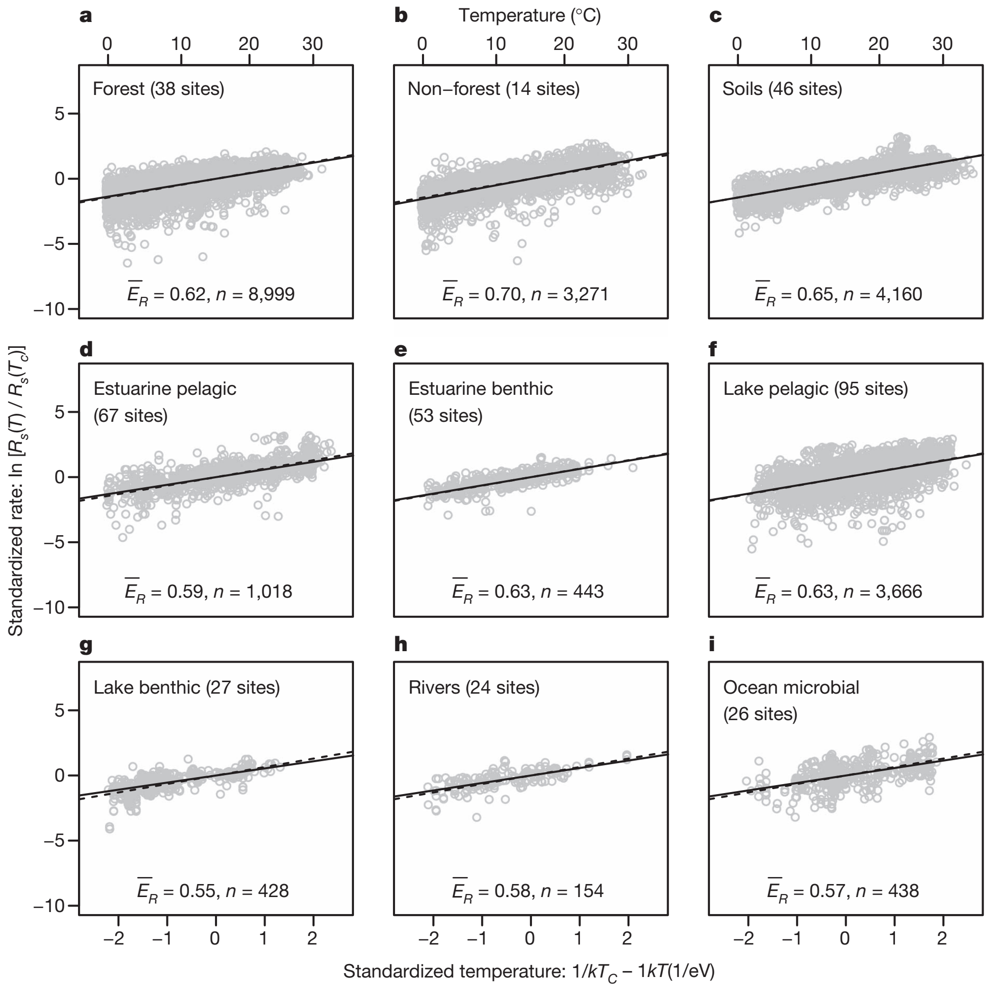


Figure 1 | Relationships between standardized short-term respiration and standardized inverse absolute temperature for 9 ecosystem types encompassing 373 sites. Standardized fluxes were expressed as $\ln[R_s(T)/R_s(T_C)]$, where $R_s(T)$ is the measured rate of respiration at site s. The site-specific estimates of respiration at fixed temperature, $R_s(T_C)$, and the average activation energies, $\bar{E}_R$, were determined using mixed-effects modelling (equation (1)). The temperature chosen for standardization ($T_C = 288\,\mathrm{K} = 15\,°\mathrm{C}$) is arbitrary and does not substantively affect our results. Standardization of the fluxes was for visualisation of the data only; $R_s(T)$ values were used in the mixed-effects analysis to determine the temperature dependencies. The solid and dashed lines correspond to the fitted and predicted ($E_R \approx 0.65$ eV) slopes, respectively. All data sets are consistent with this prediction (Table 1).

We evaluated this hypothesis using annually integrated rates of ecosystem respiration for eight ecosystem types (open oceans could not be analysed owing to insufficient long-term data; Supplementary Information 1). Across the terrestrial ecosystem types, annual respiration exhibited no significant correlation with temperature for non-forested systems (Fig. 2 and Table 1). This might be because warmer temperatures result in more negative water balances in such systems, depressing gross primary production[9,10] and thus respiration. Respiration was, however, positively correlated with temperature across forests and soils (Fig. 2 and Table 1) and, in agreement with our prediction, the temperature dependencies for the long-term rates were indistinguishable from that of photosynthesis ($E_P \approx 0.32$ eV; Fig. 2 and Table 1). We obtained similar results for these two ecosystem types after controlling for growing-season length and daylight hours during the growing season (Supplementary Information 7). In contrast, for all of the aquatic ecosystem types, annually integrated rates of respiration yielded activation energies that were consistently greater than E_P, with 95% confidence intervals (CIs) that included E_R (≈ 0.65 eV; Fig. 2 and Table 1). Overall, a likelihood ratio test indicated that the long-term temperature dependence of ecosystem respiration differed significantly ($\chi^2_1 = 17.14$, $P < 0.005$) between the terrestrial ($E = 0.32$ eV; 95% CI, 0.26–0.38 eV) and aquatic ecosystems ($E = 0.77$ eV; 95% CI, 0.57–0.98 eV), in agreement with our hypothesis.

In terrestrial ecosystems, where long-term respiration is constrained by gross primary production, equations (5) and (6) yield the prediction that the rate of ecosystem respiration at fixed temperature:

$$R(T_C) = \left(r_o^A M_A \left\langle m_A^{\alpha-1} \right\rangle + r_o^H M_H \left\langle m_H^{\alpha-1} \right\rangle\right) e^{-E_R/kT_C} \propto \left\langle e^{E_P(1kT_C - /1kT)} \right\rangle_\tau \Big/ \left\langle e^{E_R(1kT_C - /1kT)} \right\rangle_\tau \quad (7)$$

should decline with increasing average temperature according to the ratio of the factors characterizing the Boltzmann-averaged temperature kinetics for photosynthesis and respiration. This predicted decline in $R(T_C)$ with increasing temperature in part reflects a reduction in mass-corrected heterotrophic biomass[5], $M_H \langle m_H^{\alpha-1} \rangle$ (Supplementary Information 4), which occurs because respiration rates of heterotrophic organisms increase more rapidly with temperature than autochthonous net primary production[5,18,20]. Warmer ecosystems should therefore support a lower standing stock of mass-corrected heterotrophic biomass per unit of carbon fixed by photosynthesis. On the contrary, in aquatic systems, empirical data demonstrate that heterotrophic metabolism is often not directly limited by autochthonous primary production over the long term, due either to allochthonous carbon subsidies, or because primary production exceeds heterotrophic metabolism[11,15,22]. Consequently, we would expect the temperature dependence of photosynthesis to have a weak or negligible effect on

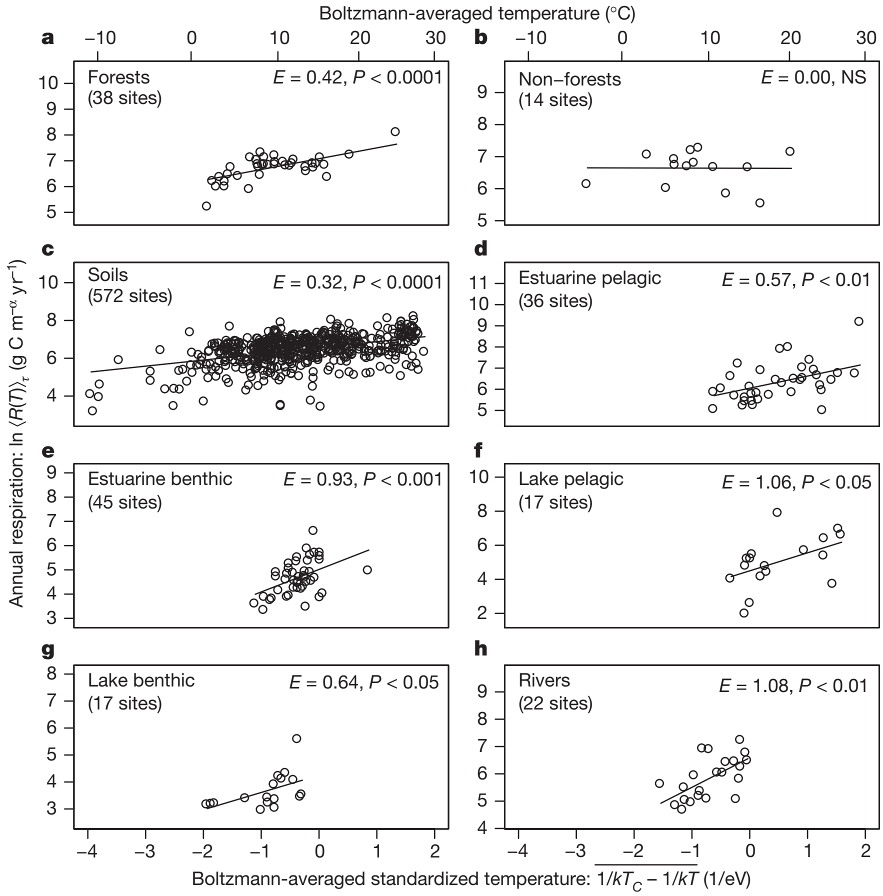


Figure 2 | Relationship between annual ecosystem respiration and Boltzmann-averaged standardized temperature. Units of fluxes are in $g\,C\,m^{-2}\,yr^{-1}$ for all ecosystem types except estuarine and lake pelagic zones, where fluxes are given in volumetric units ($g\,C\,m^{-3}\,yr^{-1}$). The long-term (that is, annual) temperature dependencies of respiration for all of the terrestrial ecosystems yielded slopes significantly less than the activation energy of respiration ($E_R \approx 0.65$ eV; Table 1), consistent with the hypothesis that photosynthesis limits heterotrophic biomass and overall respiratory metabolism (equation (6)). By contrast, for all of the aquatic ecosystems, the 95% CIs encompassed E_R. Overall, these findings suggest fundamental differences in the factors structuring aquatic and terrestrial communities along geographic temperature gradients. NS, not significant.

$R(T_C)$ (Supplementary Information 4). We evaluated these predictions using the estimates of $R(T_C)$ obtained from the mixed-effects analyses above (equation (1)). As predicted by equation (7), $R(T_C)$ was negatively correlated with average temperature for forests, non-forested ecosystems and soils (Supplementary Information 10). Thus, because annual ecosystem respiration is the product of the sizes, abundances and metabolic rates of all the organisms comprising that ecosystem, declines in $R(T_C)$ with increasing annual temperature explain how the weaker apparent temperature dependence of long-term ecosystem respiration across terrestrial systems arises. By contrast, estimates of $R(T_C)$ were not correlated with average temperature for any of the aquatic ecosystem types (Supplementary Information 10), contrary to what was observed in terrestrial ecosystems, but in line with expectations given that the short-term and long-term aquatic flux data yielded similar apparent temperature sensitivities.

Our findings highlight marked similarities in the seasonal temperature dependence of short-term respiration across the major ecosystem types on the planet (Fig. 1), as well as a clear dichotomy between terrestrial and aquatic ecosystems with respect to the temperature dependencies of annual respiration (Fig. 2) and daily respiration at fixed temperature (Supplementary Information 10). These divergent patterns between terrestrial and aquatic ecosystems can be reconciled by assuming that, over long timescales, heterotrophic metabolism in terrestrial ecosystems is more tightly constrained by autochthonous primary production[10,23], which exhibits a weaker temperature dependence than respiration (equations (6) and (7)). In many aquatic ecosystems, the constraint of autochthonous primary production on long-term respiration may be relatively weak due to a greater influence of other variables such as allochthonous carbon subsidies[11,22]. These assumptions are consistent with our compilation of long-term aquatic respiration data (with the important exception of open oceans, for which we had no long-term data), and with the observation that heterotrophic metabolism can exceed autochthonous primary production in many aquatic ecosystems over the long term[11,22]. Nevertheless, they clearly require direct empirical validation using commensurate community-level data on abundance and biomass, and ecosystem-level data on respiration and primary production. Still, our model and analyses provide a new way of reconciling the similarities and differences in the temperature dependence of ecosystem respiration across timescales and ecosystem types on the basis of the biochemical kinetics of photosynthetic and respiratory metabolisms.

METHODS SUMMARY

To estimate the seasonal temperature dependence of ecosystem respiration for each of the nine ecosystem types, linear mixed-effects models were fit to the data for each ecosystem type using the lme4 package in R statistical software (v.2.14.0)[27]. Mixed-effects modelling was also applied to all data combined to test for differences in the temperature dependence of respiration among ecosystem types. Significance was assessed using a likelihood ratio test[8], comparing a model with a common activation energy to models with different activation energies for each ecosystem type (see Supplementary Information 5).

The long-term temperature dependence of ecosystem respiration was determined by using maximum likelihood to estimate the activation energy, E, that best predicted annual ecosystem respiration based on Boltzmann-averaged temperature kinetics, $\left\langle e^{E(1/kT_C - 1/kT)} \right\rangle_\tau$ (equation (6); see Supplementary

Information 6). Boltzmann averaging is preferable to using arithmetic mean temperature to assess kinetics because this entails an approximation that becomes less accurate as seasonal variation in temperature increases[28]. Model fits are depicted graphically in Fig. 2 by re-expressing average kinetics as $\ln\left\langle e^{E(1/kT_C-1/kT)}\right\rangle_\tau\Big/E$ and $(E/k)\Big/\left[E/kT_C-\ln\left[\left\langle e^{E(1/kT_C-1/kT)}\right\rangle_\tau\right]\right]-273$ along the lower and upper horizontal axes, respectively. However, the full distributions of temperatures (including temperatures $<0\,^{\circ}$C) were used to fit the models.

Received 12 September 2011; accepted 1 May 2012.
Published online 20 June 2012.

1. Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *Science* **293,** 2248–2251 (2001).
2. Mahecha, M. D. *et al.* Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science* **329,** 838–840 (2010).
3. Gu, L., Hanson, P. J., Mac Post, W. & Liu, Q. A novel approach for identifying the true temperature sensitivity from soil respiration measurements. *Glob. Biogeochem. Cycles* **22,** GB4009 (2008).
4. Luo, Y. Terrestrial carbon-cycle feedback to climate warming. *Ann. Rev. Ecol. Evol. System.* **38,** 683–712 (2007).
5. Allen, A. P., Gillooly, J. F. & Brown, J. H. Linking the global carbon cycle to individual metabolism. *Funct. Ecol.* **19,** 202–213 (2005).
6. Cebrian, J. Patterns in the fate of production in plant communities. *Am. Nat.* **154,** 449–468 (1999).
7. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85,** 1771–1789 (2004).
8. Pinheiro, J. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer, 2000).
9. Beer, C. *et al.* Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* **329,** 834–838 (2010).
10. Raich, J. & Potter, C. S. Global patterns of carbon-dioxide emissions from soils. *Glob. Biogeochem. Cycles* **9,** 23–36 (1995).
11. del Giorgio, P. A., Cole, J., Caraco, N. & Peters, R. Linking planktonic biomass and metabolism to net gas fluxes in northern temperate lakes. *Ecology* **80,** 1422–1431 (1999).
12. del Giorgio, P. A. & Gasol, J. M. Biomass distribution in freshwater plankton communities. *Am. Nat.* **146,** 135–152 (1995).
13. Gasol, J. M. & del Giorgio, P. A. Biomass distribution in marine planktonic communities. *Limnol. Oceangr.* **42,** 1353–1363 (1997).
14. Pereira, J. S. *et al.* Net ecosystem carbon exchange in three contrasting Mediterranean ecosystems—the effect of drought. *Biogeosciences* **4,** 791–802 (2007).
15. del Giorgio, P. A. & Williams, P. J. L. B. *Respiration in Aquatic Ecosystems* (Oxford Univ. Press, 2005).
16. Enquist, B. *et al.* Scaling metabolism from organisms to ecosystems. *Nature* **423,** 639–642 (2003).
17. Dewar, R., Medlyn, B. & McMurtrie, R. Acclimation of the respiration photosynthesis ratio to temperature: insights from a model. *Glob. Change Biol.* **5,** 615–622 (1999).
18. Yvon-Durocher, G., Jones, J. I., Trimmer, M., Woodward, G. & Montoya, J. M. Warming alters the metabolic balance of ecosystems. *Philos. Trans. R. Soc. Lond.* **365,** 2117–2126 (2010).
19. Regaudie-de-Gioux, A&. Duarte, C. M. Temperature dependence of planktonic metabolism in the ocean. *Global Biogeochem. Cycles* **26,** GB1015 (2012).
20. Lopez-Urrutia, A., San Martin, E., Harris, R. P. & Irigoien, X. Scaling the metabolic balance of the oceans. *Proc. Natl Acad. Sci. USA* **103,** 8739–8744 (2006).
21. Kesselmeier, J. *et al.* Volatile organic compound emissions in relation to plant carbon fixation and the terrestrial carbon budget. *Global Biogeochem. Cycles* **16,** 1126 (2002).
22. Duarte, C. & Agusti, S. The CO_2 balance of unproductive aquatic ecosystems. *Science* **281,** 234–236 (1998).
23. Vargas, R. *et al.* Looking deeper into the soil: biophysical controls and seasonal lags of soil CO_2 production and efflux. *Ecol. Appl.* **20,** 1569–1582 (2010).
24. Chapin, S. F., Matson, P. A. & Mooney, H. A. *Principles of Terrestrial Ecosystem Ecology* (Springer, 2002).
25. Schlesinger, W. Evidence from chronosequence studies for a low carbon-storage potential of soils. *Nature* **348,** 232–234 (1990).
26. Schimel, D. *et al.* Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414,** 169–172 (2001).
27. R. Development Core Team. R: A language and environment for statistical computing http://www.R-project.org/ (2011).
28. Savage, V. H. Improved approximations to scaling relationships for species, populations, and ecosystems across latitudinal and elevational gradients. *J. Theor. Biol.* **227,** 525–534 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We wish to thank J. Cole, M. Pace, M. Reichstein and D. Baldocchi for helpful comments that greatly improved the manuscript. M. Mahecha, C. S. Hopkinson, E. Smith, C. Gudasz, C. Solomon, E. Gaiser, E. de Eyto, C.-Y. Chiu, D. Hamilton, S. Hendricks, R. Adrian, K. Rose, D. Bruesewitz, D. Richardson, M. Van de Bogert, FLUXNET and GLEON are gratefully acknowledged for supplying raw data. G.Y.-D., M.T. and G.W. acknowledge the support of the Natural Environment Research Council, UK (grant NE/F004753/1) for financial support. P.A.S. was funded by the Danish Council for Independent Research, Natural Sciences grant 10-085238 and the Danish Centre for Lake Restoration (CLEAR). J.P. acknowledges the Academy of Finland Centre of Excellence program (project number 218094) for funding. J.M.M. was supported by a Ramon y Cajal Fellowship (RYC-892 2008-03664), a Ministry of Economy grant (CGL2010-20091) and Generalitat de Catalunya grant (2009SGR142).

Author Contributions G.Y.-D. and A.P.A. analysed the data, wrote the manuscript, and devised the research. A.C., J.M.C., M.D., P.d.G., J.M.G., J.M.M., J.P., P.A.S., M.T. and G.W. commented on the manuscript. G.Y.-D., A.P.A., P.d.G., J.M.M. and M.T. discussed ideas. A.C., J.M.C., M.D., P.d.G., J.P., M.T. and P.A.S. provided raw data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.Y.-D. (g.yvon-durocher@exeter.ac.uk).

LETTER

doi:10.1038/nature11292

Inland thinning of West Antarctic Ice Sheet steered along subglacial rifts

Robert G. Bingham[1], Fausto Ferraccioli[2], Edward C. King[2], Robert D. Larter[2], Hamish D. Pritchard[2], Andrew M. Smith[2] & David G. Vaughan[2]

Current ice loss from the West Antarctic Ice Sheet (WAIS) accounts for about ten per cent of observed global sea-level rise[1]. Losses are dominated by dynamic thinning, in which forcings by oceanic or atmospheric perturbations to the ice margin lead to an accelerated thinning of ice along the coastline[2–5]. Although central to improving projections of future ice-sheet contributions to global sea-level rise, the incorporation of dynamic thinning into models has been restricted by lack of knowledge of basal topography and subglacial geology so that the rate and ultimate extent of potential WAIS retreat remains difficult to quantify. Here we report the discovery of a subglacial basin under Ferrigno Ice Stream up to 1.5 kilometres deep that connects the ice-sheet interior to the Bellingshausen Sea margin, and whose existence profoundly affects ice loss. We use a suite of ice-penetrating radar, magnetic and gravity measurements to propose a rift origin for the basin in association with the wider development of the West Antarctic rift system. The Ferrigno rift, overdeepened by glacial erosion, is a conduit which fed a major palaeo-ice stream on the adjacent continental shelf during glacial maxima[6]. The palaeo-ice stream, in turn, eroded the 'Belgica' trough, which today routes warm open-ocean water back to the ice front[7] to reinforce dynamic thinning. We show that dynamic thinning from both the Bellingshausen and Amundsen Sea region is being steered back to the ice-sheet interior along rift basins. We conclude that rift basins that cut across the WAIS margin can rapidly transmit coastally perturbed change inland, thereby promoting ice-sheet instability.

Many independent satellite sensors have been used to gauge the recent mass imbalance of the Antarctic ice sheet, thereby to assess the rate of its contribution to global sea-level rise[2–4,8]. These studies highlight coastal regions of the WAIS as major contributors to current sea-level rise through a process termed "dynamic thinning"[2]. Such a region is Pine Island Glacier in the Amundsen Sea embayment, where gradual ungrounding and thinning of the floating ice shelf, potentially triggered by oceanic forcing, have induced progressive drawdown and thinning of inland ice[3–5]. These observations have reignited concerns that the wider WAIS is progressing towards a predicted collapse within the next few centuries, increasing the need to understand the fundamental influences on its dynamic behaviour.

Theory suggests that the stability of WAIS is controlled by the subglacial topographic configuration, with landward-deepening basins favouring runaway retreat inland[9]. The subglacial geology and crustal structure of the West Antarctic rift system (WARS; Fig. 1a) may also influence ice dynamics[10–12]. Sedimentary basins underlie the onset of some ice streams in the Ross Sea embayment[13,14] and their initiation may be affected by elevated geothermal heat flux linked to inferred recent volcanism[10,15]. Yet, for the Bellingshausen Sea embayment, currently one of the most rapidly thinning parts of the WAIS (Fig. 1b), the subglacial topography and geological setting that may influence ice-sheet behaviour have remained largely unknown. Here we analyse new geophysical data sets from inland of the Bellingshausen Sea and demonstrate that propagation of ice thinning towards the interior is promoted where narrow rift basins associated with a northeasterly extension of the WARS connect to the ocean.

Our data set comprises the first systematic radar survey of Ferrigno Ice Stream (FIS; 85° W, 74° S), a 14,000-km^2 ice-drainage catchment clearly identified by satellite altimetry as the most pronounced 'hotspot' of dynamic thinning along the Bellingshausen Sea margin of the WAIS (Fig. 1b). Data were collected by over-snow survey between November 2009 and February 2010, and supplemented by airborne data collected by the US NASA Operation IceBridge programme in 2009. The only previous measurements of ice thickness across the entire 150 km × 115 km catchment were a sparse set of reconnaissance seismic and gravity spot-depths obtained 50 years previously along exploratory traverses, and a handful of airborne radar measurements collected on the way to other locations (Fig. 1c). Our new view of the bed beneath FIS reveals a narrow subglacial basin with a maximum depth of about 1,500 m below sea level striking northeast–southwest through the catchment. Viewed in the wider context (Fig. 1d), using the most recent compilations of subglacial topography and offshore bathymetry, together with new analyses of magnetic and gravity data, we can see that the basin forms part of a major fault system that connects rift basins in the interior of the WARS to Eltanin Bay[16,17] (Supplementary Fig. 1a–c).

Radar-derived bed echoes across FIS (Fig. 2a, Supplementary Figs 2–4) deviate in form from the U-shaped parabolic profile that represents the 'pure' product of glacial erosion[18]. Instead, the steeply dipping, approximately 1-km-high basin flanks and the flat basin floor resemble classical rift structures[19]. Further supporting evidence for a rift origin is the dip of the bed away from the basin flanks (Fig. 2a, Supplementary Figs 3 and 4), consistent with the tilting of fault blocks and footwall uplift.

Aeromagnetic anomaly data[20] (Fig. 3, Supplementary Figs 5–7) reveal that the Ferrigno rift formed close to the boundary between a highly magnetic magmatic arc province that lies primarily offshore in the FIS region (and that extends from the Antarctic Peninsula to Thurston Island) and a more weakly magnetic province onshore. The latter is interpreted as a back-arc region, where the early WARS developed about 105–90 million years (Myr) ago[21]. This part of the WARS was then reactivated during inferred dextral transtensional motion between 48 Myr ago and 26 Myr ago[17]. Depth to magnetic sources help to constrain our gravity and magnetic models and indicate that the narrow Ferrigno rift contains about 1 km of sedimentary infill (Fig. 3c, d, Supplementary Figs 8 and 9). Three-dimensional inversion of satellite gravity data indicates that the crust is around 25–21 km thick beneath the Ferrigno rift and the adjacent Siple Trough region (Fig. 1d and Supplementary Figs 1c and 10). This is similar to the 22–24-km-thick crust in the western Amundsen Sea embayment[22] and the approximately 20-km-thick crust beneath the Pine Island rift[23] and the Bentley Subglacial Trench[10]. These new findings lead us to conclude that the Ferrigno rift region was affected by crustal thinning associated with the WARS. There is, however, no magnetic evidence in support of widespread Cenozoic magmatism

[1]School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen, AB24 3UF, UK. [2]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK.

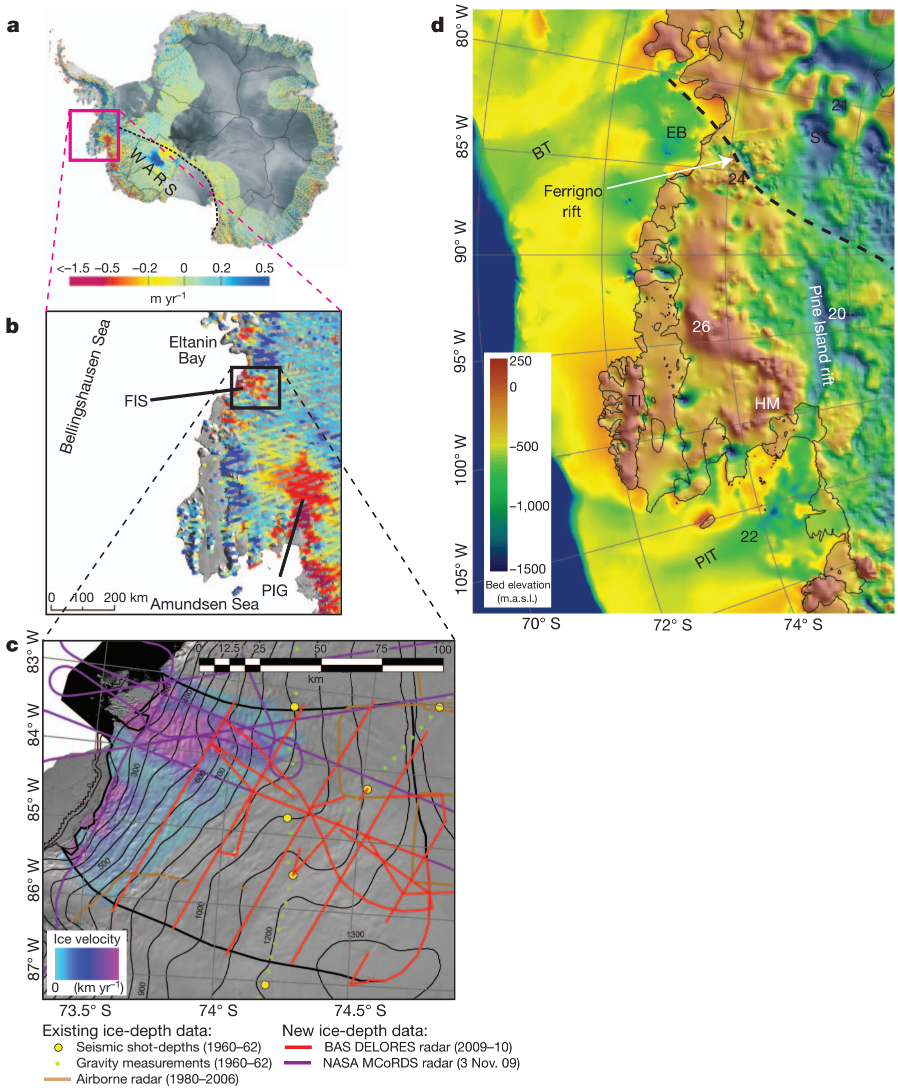


Figure 1 | Surface change, survey coverage, and subglacial topography for the Bellingshausen Sea sector of West Antarctica. **a**, Rate of surface elevation change along ICESat tracks measured between 2003 and 2007; superimposed over the Landsat Image Mosaic of Antarctica (grey shading). Regions coloured red in the Bellingshausen and Amundsen Sea sectors of West Antarctica exhibit strong ice surface lowering attributable to dynamic thinning. The flank of the WARS is annotated with a black dashed curve. **b**, Enlargement of the boxed area in **a**, with major geographical features labelled. **c**, Enlargement of the boxed area in **b**, showing orientation of new radar tracks obtained across FIS in 2009–10; pre-existing ice-depth measurement locations are also marked. The black line demarcates the Ferrigno catchment. Also shown are satellite-derived ice velocities for the lower portion only of the catchment (ref. 28), 100-m surface contours and background imagery from the Landsat Image Mosaic of Antarctica. **d**, Regional subglacial topography mosaicked from our new data set for FIS (see Fig. 3a and Supplementary Fig. 3 for detail) and refs 29 and 30. The black dashed line highlights a major fault system that we propose connects rift basins in the interior of the WARS to Eltanin Bay (EB). BT, Belgica trough; PIT, Pine Island trough; ST, Siple trough; TI, Thurston Island; HM, Hudson Mountains. Numbers are estimates of crustal thickness (in kilometres) in the WARS derived from inversion of Bouguer gravity anomaly data and seismic data[11,22]. m.a.s.l., metres above sea level.

close to the Ferrigno rift, in contrast to the Ross and Amundsen Sea sectors of the WARS (Supplementary Figs 1b and 11).

A recently suggested middle-Miocene (about 17–13 My ago) change in plate motion between East and West Antarctica[24] has significant implications for how we interpret the Ferrigno rift in relation to inferred Neogene (23–2.5 Myr ago) rifting processes (Fig. 1d and Supplementary Fig. 1). A Neogene rifting stage and glacial overdeepening may account for the deepest basins within the WARS, including the Byrd Subglacial Basin and the Bentley Subglacial Trench[17]. The low effective elastic thickness in the Pine Island rift[23] is similar to values observed over recent rifts[19], supporting the hypothesis of a Neogene rifting phase there. All these features that are inferred to be Neogene rifts may be kinematically linked (Fig. 1d and Supplementary Fig. 10).

We note that the Ferrigno rift broadly aligns with the inland extent and distribution of the dynamic-thinning signal observed at the ice surface with altimetry (Fig. 3a). Modern ice flow also follows the direction of the rift (Fig. 1c). These associations suggest that dynamic thinning of ice towards the West Antarctic interior may be promoted where ice is underlain by glacially overdeepened rift basins associated with development of the WARS.

We propose two factors that can explain the observed correlation of inland dynamic thinning with subglacial rifting. First, the sedimentary infill we identified within the Ferrigno rift (Fig. 3c, d) probably facilitated formation of a deforming bed (see ref. 14 for an example) that would promote enhanced ice flow and thinning. Second, elevated geothermal heat flux has been measured within the Terror rift (Supplementary Fig. 1a), a narrow rift basin that was active in Neogene times[25]. If the Ferrigno rift were similarly active in the Neogene, as we hypothesize, then enhanced geothermal heat flux linked to focused crustal thinning would probably occur beneath FIS. This would lead to excess generation of subglacial meltwater, which in turn would lubricate the bedrock, accelerate ice flow and thereby exacerbate dynamic thinning effects. Although incidences of rifts exploited and overdeepened by ice-stream flow have been noted elsewhere[13,26,27], in none of these cases has an influence on contemporary ice thinning been noted. What seems to be required to promote dynamic thinning inland is a coincidence of a deep rift basin cutting across the ice-sheet margin and inflow of warm ocean water onto the continental shelf.

We conclude that the WAIS is most at threat from the inward incursion of dynamic thinning along glacially overdeepened rifts that extend both into the Amundsen and Bellingshausen seas. The Ferrigno rift connects through to Eltanin Bay, where the lack of an ice shelf today and in the observational record implicates the sustained presence of relatively warm air and/or ocean water[7], providing a forcing mechanism for the dynamic thinning of coastal ice. Without favourable subglacial conditions further inland, this dynamic thinning would be limited to the coastal fringes as on other parts of the Antarctic margin; instead, the deep inland reach of the Ferrigno rift facilitates a strong coupling between the oceanic front and the deeper ice-sheet interior (Fig. 3a and Supplementary Fig. 12). The situation at the FIS evokes that of Pine Island Glacier, where, over two decades of observation, the access of warm ocean water has instigated dynamic thinning at the

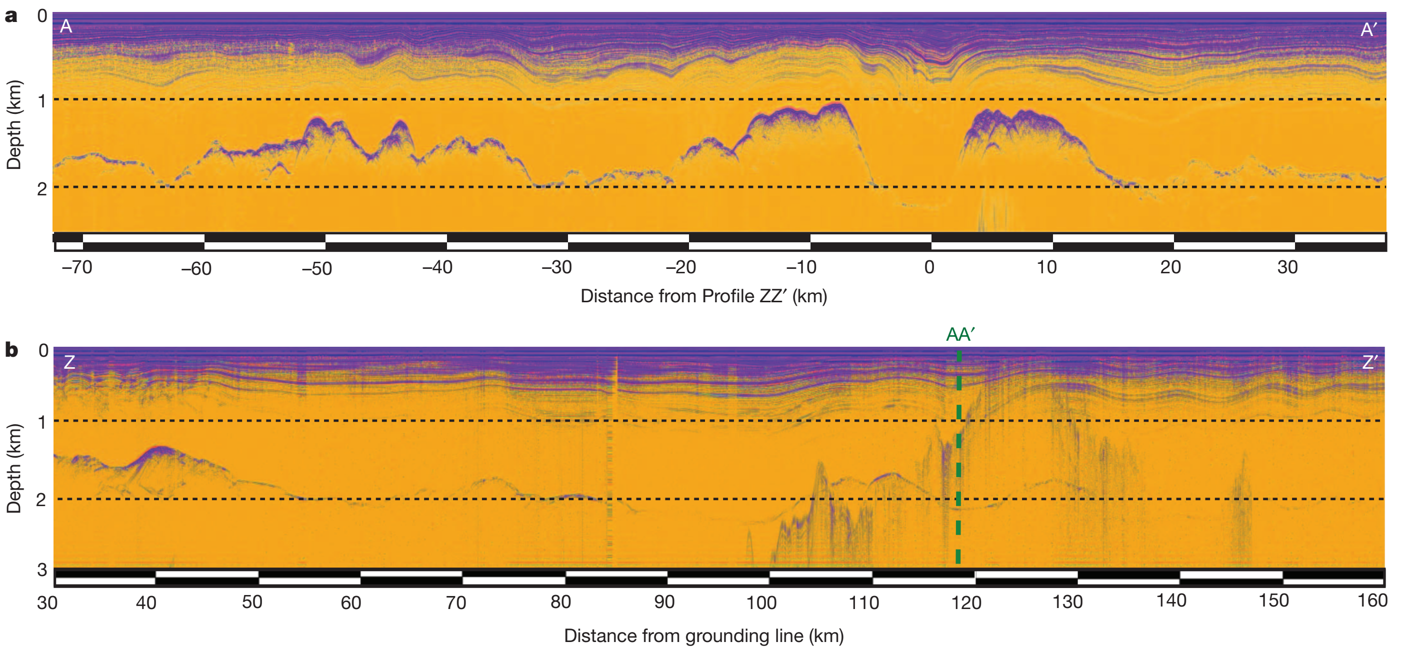


Figure 2 | Radar profiles showing morphology of Ferrigno Ice Stream bed. **a**, 110-km-long radar profile collected across the Ferrigno catchment transverse to the main axis of ice flow (profile AA′ marked on Fig. 3a). Distance scale is marked relative to the ice-stream (rift) centreline (path followed by profile ZZ′ in **b**). Ice flow is towards the reader. The vertical scales in **a** and **b** show depth below present ice surface. **b**, 130-km-long radar profile collected along the main trench ZZ′ (location marked on Fig. 3a). Distance scale is marked in kilometres from the grounding line; the crossing point of profile AA′ is shown. Sporadic interference patterns between 100 km and 147 km result from ‘sideswipe’ returns from valley walls as the radar passed over a narrowing in the trench.

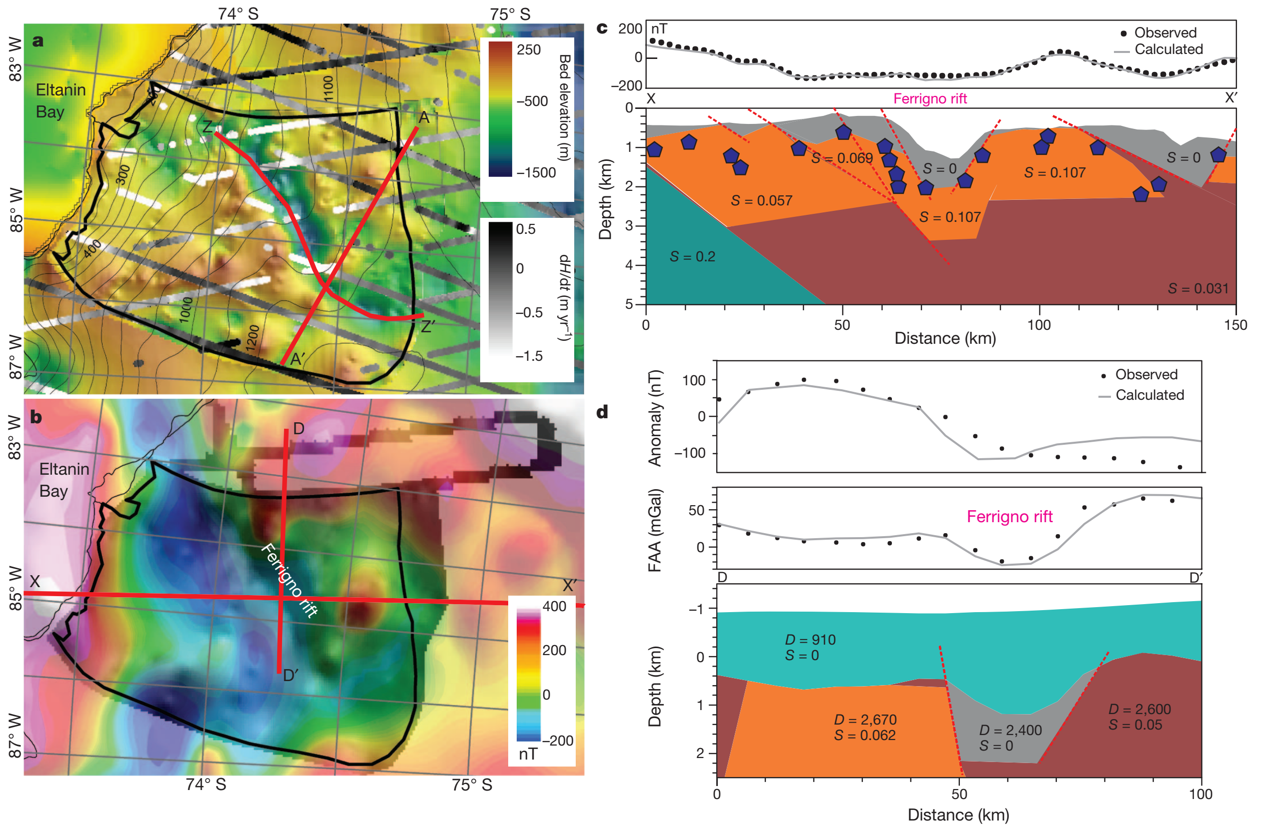


Figure 3 | Dynamic thinning of the ice sheet steered along a subglacial rift. **a**, Rate of surface ice elevation change measured between 2003 and 2007 along ICESat tracks (greyscale) superimposed over subglacial topography, showing enhanced thinning in the interior over the Ferrigno rift region. Black contours with labels show surface topography. Red lines show radar profiles in Fig. 2. **b**, The same area, reduced to the pole aeromagnetic anomaly map (colour scale, partially transparent), with new Ferrigno subglacial topography underneath. Note the northeast–southwest-oriented magnetic low that aligns with the Ferrigno rift. Red lines show the location of our magnetic and gravity models of the Ferrigno rift. **c**, Magnetic model of the Ferrigno rift. Blue symbols show the results of depth to magnetic source calculations that help constrain the model; the numbers denote magnetic susceptibilities. **d**, Combined gravity and magnetic model, indicating an approximately 1-km-thick sedimentary infill in the Ferrigno rift. Magnetic susceptibilities and densities are labelled as S (dimensionless) and D (kg m^{-3}) respectively.

grounding line, followed by propagation of ice drawdown deep along the Pine Island rift[3,23]. A radar profile we collected directly along the main axis of FIS (Fig. 2b) shows the bed deepening inland for about 70 km from the lower reaches. This leaves much of the upper catchment of FIS vulnerable to impending drawdown and retreat analogous to that of Pine Island Glacier, and demonstrates that the risk to the WAIS from dynamic thinning extends beyond the Amundsen Sea embayment.

Overall, the recent changes being witnessed today in West Antarctica represent not simply a short-term ice-sheet response to climate warming, but form part of a wider, sustained and complex system of interactions between tectonic activity, glacial landscape modification, and oceanic and atmospheric change. Over millennia, ice streams such as FIS and Pine Island Glacier have exploited rift structures and, over several glacial maxima, focused flow into palaeo-ice streams that have eroded troughs over the continental shelf such as the Belgica and Pine Island troughs (Fig. 1d). The Pine Island trough itself exploits a tectonic lineament caused by former rifting[22]; the Belgica trough also aligns with a northwest–southeast trending scarp on the inner continental shelf that runs at right angles to FIS and is probably imposed on a major tectonic lineament (Fig. 1d). These 'rift-directed' offshore troughs now form the putative routes through which warm open-ocean waters penetrate back over the continental shelf to attack the ice margin. Overdeepened rift basins onshore are now steering the transmission of this dynamic-thinning perturbation even further back towards the interior of West Antarctica, with likely consequences for ice-sheet stability.

METHODS SUMMARY

Our radar survey of FIS represents the only over-snow exploration over the region since two traverses immediately following the International Geophysical Year 50 years previously. It also comprises the first systematic survey (to our knowledge) of any of the WAIS catchments fringing the Bellingshausen Sea. Radar data were collected with the British Antarctic Survey DELORES system, a 2-MHz monopulse 'DEep-LOoking Radio Echo Sounder' with a pulse-repetition rate of 1 kHz and a digitization period of 10 ns. Tying this into dual-frequency global positioning system (GPS) for navigational fixing, we measured ice thickness at 7.5-m intervals along tracks, obtaining a total of 250,970 ice-thickness points. Assuming radar wave speed through ice of $168.5\ \mathrm{m}\ \mu\mathrm{s}^{-1}$, these were converted to bed elevations with estimated $\pm$3-m vertical resolution (see the file named ferrigno_delores.txt in the Supplementary Information). We also incorporated 17,793 further ice-thickness measurements obtained over parts of FIS on 3 November 2009 by the US NASA Operation IceBridge programme 140–230 MHz Multichannel Coherent Radar Depth Sounder (MCoRDS) (publicly available on http://nsidc.org/icebridge/portal/). We used these measurements, and the small amount of existing earlier ice-thickness measurements (http://www.antarctica.ac.uk//bas_research/data/access/bedmap/), to interpolate bed elevations over a 1-km grid mesh. We provide the final grid file used for this paper, ferrigno_topogrid_1km.txt in the Supplementary Information.

A reconnaissance aeromagnetic survey was made across the region in 1986–1987 (ref. 20), after which no further data were acquired. We reprocessed these data (see the file named mag_Drape3500_redp.XYZ in the Supplementary Information) and analysed them together with land gravity data to derive new models of the Ferrigno rift and estimate the thickness of its sedimentary infill. To assess crustal thickness beneath the WARS we applied three-dimensional inversion to satellite gravity data and airborne gravity data over the adjacent catchment of Pine Island Glacier.

Received 29 March; accepted 7 June 2012.

1. Meier, M. F. *et al.* Glaciers dominate eustatic sea-level rise in the 21st century. *Science* **317,** 1064–1067 (2007).
2. Pritchard, H. D., Arthern, R. J., Vaughan, D. G. & Edwards, L. A. Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. *Nature* **461,** 971–975 (2009).
3. Wingham, D. J., Wallis, D. W. & Shepherd, A. Spatial and temporal evolution of Pine Island Glacier thinning, 1995–2006. *Geophys. Res. Lett.* **36,** L17501 (2009).
4. Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A. & Lenaerts, J. Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophys. Res. Lett.* **38,** L05503 (2011).
5. Jacobs, S. S., Jenkins, A., Giulivi, C. F. & Dutrieux, P. Stronger ocean circulation and increased melting under Pine Island Glacier ice shelf. *Nature Geosci.* **4,** 519–523 (2011).
6. Ó Cofaigh, C. *et al.* Flow of the West Antarctic Ice Sheet on the continental margin of the Bellingshausen Sea at the Last Glacial Maximum. *J. Geophys. Res.* **110,** B11103 (2005).
7. Holland, P. R., Jenkins, A. & Holland, D. M. Ice and ocean processes in the Bellingshausen Sea, Antarctica. *J. Geophys. Res.* **115,** C05020 (2010).
8. Chen, J. L., Wilson, C. R., Blankenship, D. & Tapley, B. D. Accelerated Antarctic ice loss from satellite gravity measurements. *Nature Geosci.* **2,** 859–862 (2009).
9. Schoof, C. Ice sheet grounding line dynamics: steady states, stability, and hysteresis. *J. Geophys. Res.* **112,** F03S28 (2007).
10. Blankenship, D. D. *et al.* in *The West Antarctic Ice Sheet: Behaviour and Environment* (eds Alley, R. B. & Bindschadler, R. A.) *Antarct. Res. Ser.*, **77,** 105–121 (AGU, 2001).
11. Winberry, J. P. & Anandakrishnan, S. Crustal structure of the West Antarctic Rift System and Marie Byrd Land hotspot. *Geology* **32,** 977–980 (2004).
12. Dalziel, I. W. D. On the extent of the active West Antarctic Rift System. *Terra Antarctica Rep.* **12,** 193–202 (2006).
13. Anandakrishnan, S., Blankenship, D. D., Alley, R. B. & Stoffa, P. L. Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations. *Nature* **394,** 62–65 (1998).
14. Studinger, M. *et al.* Subglacial sediments: a regional geological template for ice flow in West Antarctica. *Geophys. Res. Lett.* **28,** 3493–3496 (2001).
15. Maule, C. F., Purucker, M. E., Olsen, N. & Mosegard, K. Heat flux anomalies in Antarctica revealed by satellite magnetic data. *Science* **309,** 464–467 (2005).
16. Müller, R. D., Gohl, K., Cande, S. C., Goncharov, A. & Golynsky, A. V. Eocene to Miocene geometry of the West Antarctic Rift System. *Aust. J. Earth Sci.* **54,** 1033–1045 (2007).
17. LeMasurier, W. E. Neogene extension and basin overdeepening in the West Antarctic rift inferred from comparisons with the East African rift and other analogs. *Geology* **36,** 247–250 (2008).
18. Harbor, J. M. Numerical modeling of the development of U-shaped valleys by glacial erosion. *Geol. Soc. Am. Bull.* **104,** 1364–1375 (1992).
19. Olsen, K. H. & Morgan, P. in *Continental Rifts: Evolution, Structure, Tectonics* (ed. Olsen, K. H.) 3–26 (Elsevier, 2006).
20. Maslanyj, M. P. & Storey, B. C. Regional aeromagnetic anomalies in Ellsworth Land: crustal structure and Mesozoic microplate boundaries within West Antarctica. *Tectonics* **9,** 1515–1532 (1990).
21. Siddoway, C. S. in *Antarctica: a Keystone in a Changing World* (eds Cooper, A. K. *et al.*), *Proc. 10th Int. Symp. Ant. Sci.* 91–114 (The National Academies Press, 2008).
22. Gohl, K. Basement control on past ice sheet dynamics in the Amundsen Sea Embayment, West Antarctica. *Palaeogeogr. Palaeoclim. Palaeoecol.* **335/ 6,** 35–41 (2012).
23. Jordan, T. A. *et al.* Aerogravity evidence for major crustal thinning under the Pine Island Glacier region (West Antarctica). *Geol. Soc. Am. Bull.* **122,** 714–726 (2010).
24. Granot, R., Cande, S. C., Stock, J. M., Davey, F. J. & Clayton, R. W. Postspreading rifting in the Adare Basin, Antarctica: regional tectonic consequences. *Geochem. Geophys. Geosyst.* **11,** Q08005 (2010).
25. Schröder, H., Paulsen, T. & Wonik, T. Thermal properties of the AND-2A borehole in the southern Victoria Land Basin, McMurdo Sound, Antarctica. *Geosphere* **7,** 1324–1330 (2011).
26. Taylor, J. *et al.* Topographic controls on post-Oligocene changes in ice-sheet dynamics, Prydz Bay region, East Antarctica. *Geology* **32,** 197–200 (2004).
27. De Angelis, H. & Kleman, J. Palaeo-ice streams in the Foxe/Baffin sector of the Laurentide Ice Sheet. *Quat. Sci. Rev.* **26,** 1313–1331 (2007).
28. Rignot, E. *et al.* Recent Antarctic ice mass loss from radar interferometry and regional climate modelling. *Nature Geosci.* **1,** 106–110 (2008).
29. Vaughan, D. G. *et al.* New boundary conditions for the West Antarctic ice sheet: subglacial topography beneath Pine Island Glacier. *Geophys. Res. Lett.* **33,** L09501 (2006).
30. Graham, A. G. C., Nitsche, F. O. & Larter, R. D. An improved bathymetry compilation for the Bellingshausen Sea, Antarctica, to inform ice-sheet and ocean models. *Cryosphere* **5,** 95–106 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by the Natural Environment Research Council (NERC/AFI/CGS/11/60) and British Antarctic Survey research programme Polar Science for Planet Earth. We acknowledge NASA Operation IceBridge for airborne ice-sounding data, A. G. C. Graham for bathymetry data and C. Griffiths for field assistance.

Author Contributions All authors contributed to research design. R.G.B. performed the field research. R.G.B. and E.C.K. processed radar data. F.F. analysed and interpreted the aeromagnetic and aerogravity data; all authors participated in data discussion and interpretation; R.G.B. and F.F. wrote the manuscript; and all authors contributed substantial comments and editorial revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.G.B. (r.bingham@abdn.ac.uk).

doi:10.1038/nature11228

ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation

Tatsuo Hashimoto[1,2]*, Thomas Perlot[1]*, Ateequr Rehman[3]†, Jean Trichereau[1], Hiroaki Ishiguro[2], Magdalena Paolino[1], Verena Sigl[1], Toshikatsu Hanada[1], Reiko Hanada[1], Simone Lipinski[3], Birgit Wild[4], Simone M. R. Camargo[5], Dustin Singer[5], Andreas Richter[4], Keiji Kuba[6], Akiyoshi Fukamizu[7], Stefan Schreiber[3], Hans Clevers[8], Francois Verrey[5], Philip Rosenstiel[3] & Josef M. Penninger[1]

Malnutrition affects up to one billion people in the world and is a major cause of mortality[1,2]. In many cases, malnutrition is associated with diarrhoea and intestinal inflammation, further contributing to morbidity and death[2]. The mechanisms by which unbalanced dietary nutrients affect intestinal homeostasis are largely unknown. Here we report that deficiency in murine angiotensin I converting enzyme (peptidyl-dipeptidase A) 2 (*Ace2*), which encodes a key regulatory enzyme of the renin-angiotensin system (RAS), results in highly increased susceptibility to intestinal inflammation induced by epithelial damage. The RAS is known to be involved in acute lung failure[3], cardiovascular functions[4] and SARS infections[5]. Mechanistically, ACE2 has a RAS-independent function, regulating intestinal amino acid homeostasis, expression of antimicrobial peptides, and the ecology of the gut microbiome. Transplantation of the altered microbiota from *Ace2* mutant mice into germ-free wild-type hosts was able to transmit the increased propensity to develop severe colitis. ACE2-dependent changes in epithelial immunity and the gut microbiota can be directly regulated by the dietary amino acid tryptophan. Our results identify ACE2 as a key regulator of dietary amino acid homeostasis, innate immunity, gut microbial ecology, and transmissible susceptibility to colitis. These results provide a molecular explanation for how amino acid malnutrition can cause intestinal inflammation and diarrhoea.

The RAS has now been studied for more than a century. It is regulated by the opposing actions of two key carboxypeptidases, angiotensin converting enzyme (ACE) and ACE2 (refs 4, 6, 7). In addition to its catalytic activity, we and others have recently found that ACE2 associates with the neutral amino acid transporter B^0AT1 and that ACE2 is required for expression of this transporter on the luminal surface of intestinal epithelial cells[8,9]. Variants in the gene that encodes B^0AT1 (*SLC6A19*) have been identified as a cause of Hartnup's disease[10], a rare autosomal recessive disorder associated with pellagra-like symptoms including diarrhoea that manifests under conditions of malnutrition[11]. However, the *in vivo* function of ACE2 in the gut epithelium remained to be investigated.

Morphological and ultra-structural analysis of the small and large intestine of *Ace2* knockout mice (see below) did not reveal any alterations. When we challenged such mice with dextran sodium sulphate (DSS), an irritant that disrupts the intestinal epithelial barrier and results in colitis[12], a profoundly increased inflammatory reaction was observable compared to wild-type littermates. We detected an enhanced infiltration of inflammatory cells, significant shortening of the colon length, increased intestinal bleeding, crypt damage, weight loss and severe diarrhoea, resulting in an overall increased disease activity[13] (Fig. 1a–c, Supplementary Fig. 1). Similar results were obtained using trinitrobenzene sulphonic acid (TNBS)-induced colitis (Supplementary Fig. 2). Thus, genetic inactivation of the key RAS enzyme ACE2 results in severe colitis following intestinal injury. We note that *Ace2* deficiency was investigated in male mice ($Ace2^{-/y}$, the *Ace2* gene being located on the X chromosome) throughout this study, however, we observed a similar phenotype in female $Ace2^{-/-}$ mice (not shown).

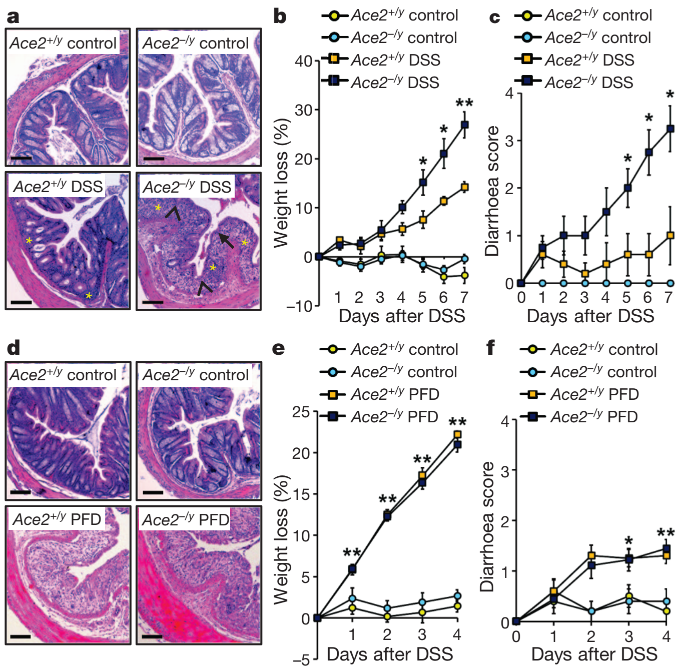


Figure 1 | *Ace2* deficiency and protein malnutrition worsen DSS-induced colitis. **a**, Colon histopathology, **b**, percentage weight loss, and **c**, diarrhoea scores in control and DSS-treated $Ace2^{+/y}$ and $Ace2^{-/y}$ littermates. In **a**, note crypt damage (arrowheads), ulcerations (arrow), and infiltration of inflammatory cells (asterisks) in DSS-treated $Ace2^{-/y}$ mice. Haematoxylin and eosin staining on day 7 after DSS challenge. Scale bars, 100 μm. **d**, Colon histopathology (haematoxylin and eosin staining, day 4 after DSS challenge; scale bars, 100 μm), **e**, percentage weight loss, and **f**, diarrhoea scores of DSS-treated $Ace2^{+/y}$ and $Ace2^{-/y}$ littermates fed either normal chow (Control) or a protein free diet (PFD; <0.2% protein). All values are mean ± s.e.m. of 5–9 mice per group. $*P < 0.05$, $**P < 0.01$ comparing DSS-treated $Ace2^{+/y}$ with $Ace2^{-/y}$ littermates, or $Ace2^{+/y}$ mice on normal diet with those on PFD (paired-*t*-test).

[1]IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, 1030 Vienna, Austria. [2]The Department of Medical Science and Cardiorenal Medicine, Yokohama City University Graduate School of Medicine and School of Medicine, 2360004 Yokohama, Japan. [3]Institute of Clinical Molecular Biology, University of Kiel, Schittenhelmstrasse12, 24105 Kiel, Germany. [4]Department of Chemical Ecology and Ecosystem Research, Center of Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. [5]Institute of Physiology and Center for Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. [6]Department of Biological Informatics and Experimental Therapeutics, Akita University Graduate School of Medicine, 1-1-1 Hondo, Akita 010-8543, Japan. [7]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. [8]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences and University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands. †Present address: Department of Environmental Health Sciences, University Medical Center, Breisacher Strasse 115b, D-79106 Freiburg, Germany.
*These authors contributed equally to this work.

To test whether the severe colitis phenotype is due to the catalytic activity of ACE2, we treated mice with recombinant soluble ACE2 (rsACE2). We and others have shown that rsACE2 rescues virtually all previously reported *in vivo* ACE2 functions within the RAS system[3,14]. As it lacks the transmembrane domain, soluble ACE2 cannot associate with the neutral amino acid transporter B^0AT1 and thus cannot stabilize its cell-surface localization[8]. Treatment of mice with rsACE2 did not rescue the severe DSS-induced colitis phenotype of *Ace2* mutant mice (Supplementary Fig. 3a–h). One *in vivo* consequence of *Ace2* deletion within the RAS is accumulation of angiotensin II (ANGII), which then primarily acts on the G-protein-coupled AT1a receptor (Agtr1a)[3]. We indeed observed increased AngII levels in the colon of DSS-treated *Ace2* mutant mice that were reduced to background levels after treatment with rsACE2 (Supplementary Fig. 3i). To rule out the possibility that locally increased AngII might cause severe colitis, we generated *Agtr1a*$^{-/-}$*Ace2*$^{-/y}$ mice. Deletion of *Agtr1* did not rescue the severe DSS-induced colitis (Supplementary Fig. 4), indicating that these effects are independent of the classical RAS system. In addition to cleaving AngII, ACE2 exhibits catalytic activity towards a second peptide system, Apelin[15]. However, DSS-induced colitis was not altered in mice carrying genetic mutations in *Apelin* (Supplementary Fig. 5) or its receptor *Apj* (Supplementary Fig. 6). Thus, the catalytic activity of ACE2, essential for its function in the RAS and Apelin cleavage, has no overt role in DSS-induced intestinal inflammation.

It had been reported that the RAS can control immune functions[16]. However, in unchallenged *Ace2* mutant mice, we did not observe any apparent differences in immune cell populations of the colon and small intestine (not shown). TLR1–9 mRNA levels in jejunum, ileum, and colon were largely similar in *Ace2* deficient and wild-type mice, with the exception of TLR5 and TLR9 in the colon (Supplementary Fig. 7). Transplantation of bone marrow from *Ace2* mutant mice (knock-out, KO) into wild-type (WT) mice (KO→WT) did not result in any obvious effect on DSS-induced colitis as compared to WT→WT transplant recipients. By contrast, reconstitution of *Ace2* mutant mice with WT bone-marrow (WT→KO) led to an increased severity of DSS-induced colitis that was indistinguishable from the KO→KO control group (Supplementary Fig. 8). Thus, ACE2 exerts its role on intestinal inflammation in non-haematopoietic cells.

ACE2 expression was primarily mapped to the luminal surface of differentiated small intestinal epithelial cells, whereas lower ACE2 expression is observed in crypt cells and the colon (Supplementary Figs 9a, b and 10a). *Ace2* inactivation had no overt effect on basal proliferation of small intestinal or colon crypt cells (Supplementary Fig. 10a, b) nor did it affect apoptosis rates of intestinal epithelial cells (Supplementary Fig. 10a, c). ACE levels were slightly, albeit not significantly, increased in the jejunum of *Ace2*$^{-/y}$ mice (Supplementary Fig. 9c, d). As reported previously[8], protein expression of the neutral amino acid transporter B^0AT1 was absent in the small intestine of *Ace2* mutant mice, whereas mRNA expression was not affected (Supplementary Figs 9e, f and 11a). Owing to the lack of intestinal B^0AT1 protein expression, serum levels of the neutral amino acids valine (Val), threonine (Thr) and tyrosine (Tyr), and the essential amino acid tryptophan (Trp) were markedly reduced in *Ace2*$^{-/y}$ mice (Supplementary Fig. 11b). The marked reduction in tryptophan correlates with human data, as patients suffering from Hartnup's disease caused by *Slc6a19* mutations or malnutrition such as in anorexia nervosa exhibit reduced tryptophan serum levels[17,18].

ACE2 is a chimaeric protein that emerged from the duplication and fusion of two genes: it is homologous with ACE at the catalytic domain and with collectrin (TMEM27) in the membrane proximal domain[19]. Inactivation of collectrin in mice (*Tmem27*$^{-/-}$) results in a near complete downregulation of apical amino acid transporters such as B^0AT1 in the kidney, thereby regulating renal amino acid re-absorption[19]. By contrast, ACE2 associates with B^0AT1 on the luminal surface of intestinal epithelial cells, that is, ACE2 and B^0AT1 regulate the uptake of neutral amino acids in the intestine where collectrin is not expressed (Supplementary Figs 11a, c and 12). In both *Tmem27*$^{-/-}$ and *Ace2*$^{-/y}$ mutant mice, tryptophan is reduced in the serum (Supplementary Fig 11b, and ref. 19). To test whether impaired re-absorption of amino acids in the kidney might also result in altered susceptibility to colitis, we challenged *Tmem27*$^{-/-}$ mice with DSS. However, genetic ablation of *Tmem27* did not affect DSS-induced colitis (Supplementary Fig. 13). Although genetic inactivation of both *Ace2* or *Tmem27* result in a similar deregulation of amino acids in the serum, functionally only the loss of ACE2 affects the susceptibility to intestinal inflammation.

We, therefore, speculated that the function of ACE2 in intestinal uptake of dietary amino acids might underlie the observed phenotype. To test this hypothesis, we fed *Ace2* mutant and control mice a protein free-diet (PFD) to eliminate dietary amino acids and challenged these mice with DSS at a dose (1%) that triggers only very mild colitis. Basal weight loss was comparable among control and *Ace2* mutant mice during the period of the PFD (Supplementary Fig. 14a). Moreover, average food intake did not show a significant difference between *Ace2* deficient and control mice fed PFD or normal chow (Supplementary Fig. 14b). Importantly, PFD markedly worsened DSS-triggered colitis in wild-type mice to levels seen in *Ace2* mutants (Fig. 1d–f; Supplementary Fig. 14c–g). Thus, protein malnutrition alters the severity of DSS-induced intestinal inflammation.

The general lack of dietary protein might induce broad effects on several organ systems. We, therefore, set out to define whether a specific amino acid might be responsible for the severe inflammatory phenotype. Because serum tryptophan was markedly decreased in *Ace2* mutant mice, we focused on this essential amino acid.

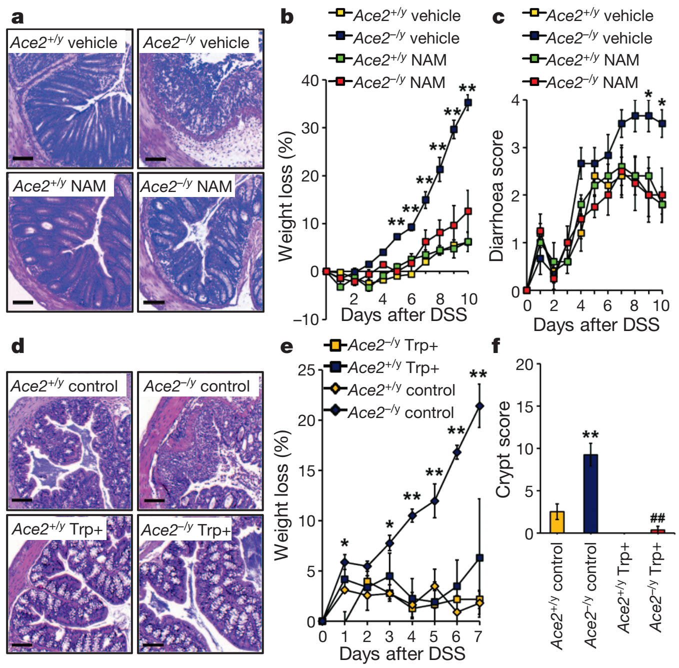


Figure 2 | Rescue of severe colitis with nicotinamide or tryptophan dipeptides. **a**, Colon histopathology (haematoxylin and eosin, day 10 after DSS challenge; scale bars, 100 μm), **b**, percentage weight loss, and **c**, diarrhoea scores of DSS-treated *Ace2*$^{+/y}$ and *Ace2*$^{-/y}$ littermates that received vehicle or nicotinamide (NAM) in their drinking water. Nicotinamide treatment was started 3 days before DSS challenge. **d**, Colon histopathology (haematoxylin and eosin, day 7; scale bars, 100 μm), **e**, percentage weight loss, and **f**, crypt injury scores of *Ace2*$^{+/y}$ and *Ace2*$^{-/y}$ mice fed a di-peptidic tryptophan diet (Trp+) or normal chow (Control). Values are mean ± s.e.m. of 3–10 mice per group. $*P < 0.05$, $**P < 0.01$ comparing *Ace2*$^{-/y}$ mice on a normal diet with those on Trp+ diet, or vehicle- versus nicotinamide-treated *Ace2*$^{-/y}$ mice. $\#\#P < 0.01$ comparing *Ace2*$^{+/y}$ versus *Ace2*$^{-/y}$ mice (paired-*t*-test).

Tryptophan is required for the *in vivo* generation of nicotinamide (also known as vitamin B3 or niacin)[17], and insufficient niacin or tryptophan in the diet is the cause of pellagra, a disease still endemic in many countries with protein malnutrition[20]. More than 90% of pellagra patients develop colitis[11] and for nearly 80 years nicotinamide has been used as treatment for pellagra. In *Ace2* mutant mice, nicotinamide almost completely alleviated the severe colitis and diarrhoea (Fig. 2a-c; Supplementary Fig. 15). To demonstrate that the severe colitis in *Ace2* deficient mice is due to impaired tryptophan uptake, we provided tryptophan in form of a Gly-Trp dipeptide to bypass the loss of the single amino acid transporter B^0AT1. Dietary dipeptidic tryptophan (a Trp+ diet) restored serum tryptophan levels (Supplementary Fig. 16a) and rescued the enhanced DSS-susceptibility of $Ace2^{-/y}$ mice (Fig. 2d–f; Supplementary Fig. 16b–d). A tryptophan-free (Trp−) diet resulted in a very marked increase in susceptibility to DSS-induced inflammation (Supplementary Fig. 17a–h). In addition, a Trp− diet markedly worsened weight loss in wild-type mice challenged with TNBS (Supplementary Fig. 17i–l). Our results show that deficiency of *Ace2* causes a critical impairment of local tryptophan homeostasis which alters the susceptibility to intestinal inflammation.

It is conceivable that an altered amino acid availability in the epithelium could lead to decreased regenerative responses and repair mechanisms after epithelial injury[21,22]. However, proliferation rates and cell death of small intestinal and colon epithelial cells were comparable between control and *Ace2* mutant mice following DSS challenge (Supplementary Fig. 18). However, expression of multiple antimicrobial peptides was markedly reduced in isolated gut ileal epithelial cells from unchallenged *Ace2* mutant mice (Fig. 3a; Supplementary Fig. 19). A Trp− diet for wild-type mice also resulted in a significant downregulation of antimicrobial peptides (Fig. 3b). Similar downregulation of antimicrobial peptides was observed on feeding wild-type mice a PFD (Supplementary Fig. 20a). By contrast, dietary nicotinamide (Supplementary Fig. 20b–g) and a Trp+ diet (Fig. 3c and Supplementary Fig. 20h, i) triggered induction of antimicrobial peptides in intestinal epithelial cells from *Ace2* mutant and wild-type mice. Thus, dietary tryptophan controls expression of small intestinal antimicrobial peptides.

Amino acids and nicotinamide can activate mTOR, which is involved in cell proliferation, survival, protein synthesis and transcription[23]. In epithelial intestinal cells from unchallenged *Ace2* mutant mice, we observed markedly impaired p70S6 kinase activity (Supplementary Fig. 21a) as well as reduced S6 phosphorylation (Fig. 3d and Supplementary Fig. 21b), indicative of reduced mTOR activity. A similar reduction of mTOR activity was recently reported in mice lacking B^0AT1 (ref. 24). Administration of a Trp+ diet resulted in increased mTOR activity in the small intestine of *Ace2* mutant mice (Fig. 3e). *In vivo* inhibition of mTOR with rapamycin in wild-type mice resulted in a significant downregulation of antimicrobial peptide expression, an effect that could not be rescued by nicotinamide (Supplementary Fig. 22a). Moreover, administration of rapamycin before the first challenge with DSS increased the severity of colitis (Fig. 3f; Supplementary Fig. 22b–h). Notably, although acute mTOR inhibition may have beneficial effects in murine colitis models[25], mTOR blockade has not been proven successful in human clinical trials in inflammatory bowel disease[26]. Thus, at the molecular level, nicotinamide and dietary tryptophan appear to exert their effects on intestinal antimicrobial peptides and colitis via the mTOR pathway.

How does ACE2 regulated uptake of tryptophan in the small intestine affect DSS-induced inflammation in the colon? It has been shown that α-defensins secreted into the ileum lumen persist in a functional state throughout the entire gut[27] and that alterations in antimicrobial peptides can affect the ecology of the small and large bowel microbiota[28]. We therefore performed deep profiling of the intestinal microbiome of *Ace2* mutant and wild-type littermates using 16S rDNA fingerprinting. The luminal ileocaecal microbiome of *Ace2* mutant mice was markedly altered, as shown by weighted (Bray–Curtis) (Fig. 4a) and unweighted (Unifrac) (Fig. 4b) analyses. Distinct operational taxonomical units (OTUs) are overrepresented in *Ace2* mutant mice (Fig. 4c; Supplementary Tables 1 and 2). Rapamycin treatment resulted in a distinct alteration of the ileocaecal gut microbiome in wild-type animals; but the altered microbiome was more closely related to that found in untreated wild-type animals than that found in untreated $Ace2^{-/y}$ animals (Supplementary Fig. 23; Supplementary Tables 3 and 4). Rapamycin acts on many different cell types that in addition to the changes in intestinal epithelial cells could alter microbiota composition[29]. Importantly, both Trp+ diet and nicotinamide treatment reverted the composition of the intestinal microbiota of *Ace2* mutant mice to be more similar to that of untreated wild-type littermates (Fig. 4a–c, Supplementary Fig. 24, Supplementary Tables 1, 2, 5 and 6). Thus, de-regulation of tryptophan amino acid homeostasis in *Ace2* mutant mice alters the intestinal microbiome.

Consistent with an alteration in the intestinal microbiome, antibiotic treatment of *Ace2* mutant mice alleviated the severe colitis (Supplementary Fig. 25). Finally, we performed gut microbiome transplants from control and *Ace2* mutant mice into germ-free hosts followed by DSS challenge. Transplantation of the ileocaecal gut microbiota from *Ace2* mutant mice transmitted the inflammatory phenotype to wild-type hosts, which was not the case for mice that were colonized with the microbiota of $Ace2^{+/y}$ mice (Fig. 4d, e; Supplementary Fig. 26a–f). Expression of antimicrobial peptides in wild-type mice did not change significantly upon transplantation of $Ace2^{+/y}$ or $Ace2^{-/y}$ microbiota within the timeframe of analysis (Supplementary Fig. 26g).

We propose the following scenario for how malnutrition and tryptophan deficiency can cause diarrhoea and colitis. Dietary tryptophan is primarily absorbed via the B^0AT1/ACE2 transport pathway on the luminal surface of small intestinal epithelial cells. This results in the activation of mTOR, either directly through nutrient sensing and/or through the tryptophan-nicotinamide pathway.

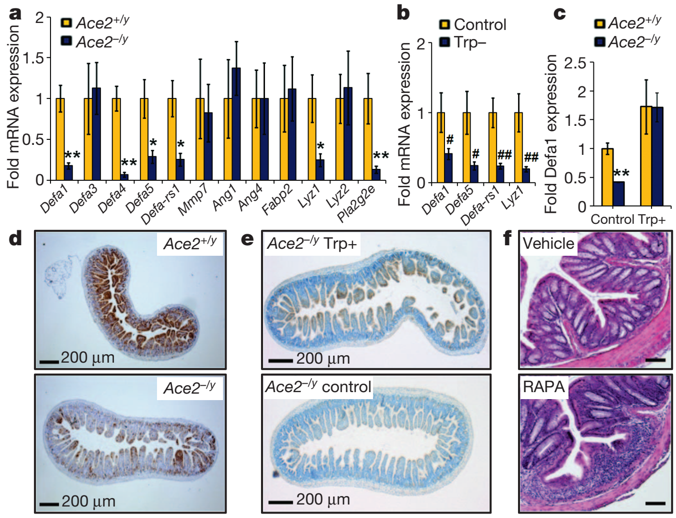


Figure 3 | Tryptophan controls antimicrobial peptides and mTOR activity. **a**, **b**, mRNA expression levels of antimicrobial peptides in epithelial cells isolated from the small intestine of **a**, unchallenged $Ace2^{+/y}$ and $Ace2^{-/y}$ littermates, and **b**, $Ace2^{+/y}$ mice fed a tryptophan-free diet (Trp−) or normal chow (Control). **c**, mRNA expression levels of antimicrobial peptide Defa1 in $Ace2^{+/y}$ and $Ace2^{-/y}$ littermates fed a Trp+ diet or normal chow (Control) for 10 days. **d**, **e**, Immunohistochemistry to detect levels of phosphorylated S6 (brown) in the small intestine of **d**, unchallenged $Ace2^{+/y}$ and $Ace2^{-/y}$ littermates or **e**, $Ace2^{-/y}$ mice fed a Trp+ or normal chow diet (Control). Scale bars, 200 μm. **f**, Colon histopathology (haematoxylin and eosin, day 8; scale bars, 100 μm) of DSS treated wild-type mice receiving vehicle or rapamycin (RAPA) i.p., initiated 6 days before DSS challenge. Values are mean ± s.e.m. of 5–6 mice per group. $*P < 0.05$, $**P < 0.01$ comparing $Ace2^{+/y}$ with $Ace2^{-/y}$ mice; $^{\#}P < 0.05$, $^{\#\#}P < 0.01$ comparing $Ace2^{+/y}$ mice on normal diet with those on Trp− diet (paired-*t*-test).

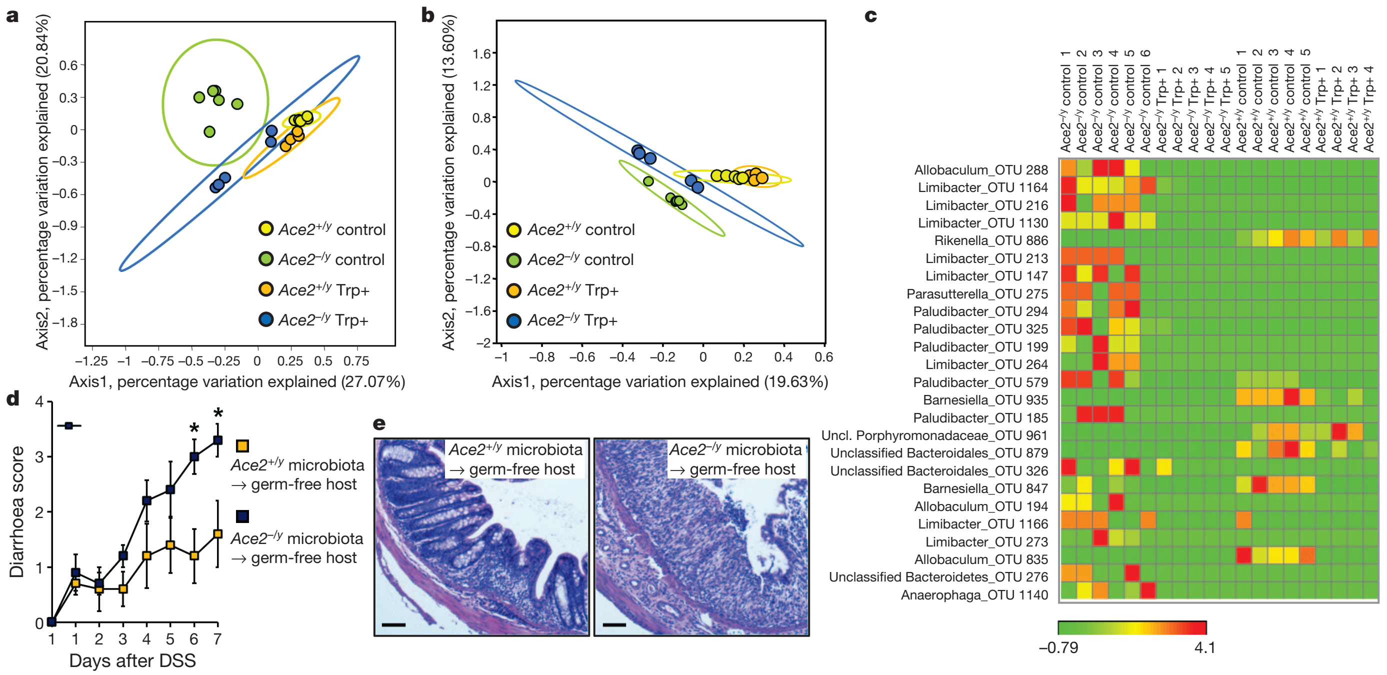


Figure 4 | Altered gut bacteria from *Ace2* mutant mice can confer susceptibility to colitis. **a**, **b**, Principal coordinate analysis plots; **a**, calculated by Bray–Curtis algorithm and **b**, based on unweighted UniFrac analysis. Plots show the similarity among ileocaecal bacterial communities in $Ace2^{+/y}$ and $Ace2^{-/y}$ mice fed a Trp+ diet or normal chow (Control) for 10 days. Only the two axes with high R^2 values are shown (axis 1, $R^2 = 0.335$; axis 2, $R^2 = 0.8116$). Each dot represents data from an individual animal. **c**, Comparison of microbial communities in $Ace2^{+/y}$ and $Ace2^{-/y}$ mice fed a Trp+ diet or normal chow (Control). The heat map depicts abundance of the top 25 species level OTUs contributing significantly to the axis shown in the weighted principal coordinate analysis plot (**a**). **d**, Diarrhoea scores and **e**, colon histopathology (haematoxylin and eosin, day 7; scale bars, 100 μm) of DSS challenged germ-free mice that received intestinal microbiota from $Ace2^{+/y}$ or $Ace2^{-/y}$ littermates. Values are mean ± s.e.m. of 4–6 mice per group. *$P < 0.05$ (paired-*t*-test).

mTOR then regulates expression of antimicrobial peptides, which affect the intestinal composition of the gut microbiota. Whether pathways other than mTOR and antimicrobial peptides also contribute to the altered microbiome needs to be explored. Our results further indicate that under conditions of intestinal injury, such an altered microbiome contributes to the severity of colitis. Importantly, our data indicate that the essential amino acid tryptophan and its metabolite nicotinamide are key regulators of gut microbiota and of propensity to inflammation.

Malnutrition is a major global health burden, affecting up to one billion people in the world and its consequences such as colitis and diarrhoea are often fatal[1,2]. In addition, chronic inflammatory conditions of the intestine, cachexia, anorexia nervosa, or amino acid malabsorption constitute severe clinical problems[30]. Our results provide novel mechanistic insights into how protein malnutrition can lead to colitis and diarrhoea. Our data also identify a molecular crosstalk between the RAS system and intestinal amino acid homeostasis via ACE2—and provide a direct link between dietary amino acid metabolism and innate immunity, the composition of the intestinal microbiota, and susceptibility to colitis.

METHODS SUMMARY

Colitis models. *Ace2*, *Tmem27*, *Agtr1*, *Apj* and *Apelin* mutant mice have been described. Mice were given protein-free and tryptophan-free diet, tryptophan dipeptides, nicotinamide, rapamycin, or were treated with catalytically active murine ACE2 protein and challenged with DSS or TNBS. Bone marrow transplantations were performed in lethally irradiated *Ace2* mutant and wild-type mice. For microbiota transplantation, caecal and ileum extracts from $Ace2^{+/y}$ or $Ace2^{-/y}$ mice were transplanted to germ-free hosts. Severity of colitis was monitored as described[13]. All experiments were performed in accordance with institutional guidelines.

Expression analyses. Small and large intestinal epithelial cells and mucosal cells were purified and gene expression assessed using quantitative RT–PCR. Protein expression was determined by western blot and localized using immunohistochemistry. Immune cell populations were determined by FACS.

Microbiota sequencing. Genomic DNA was extracted from the terminal ileum, amplified by PCR, and sequenced using Roche 454 Titanium chemistry. All sequence reads were filtered for quality and length using PANGEA. Sequences were clustered using average neighbour algorithm into species level operational taxonomical units (OTUs). Distance matrices between samples were generated based on weighted (Bray–Curtis similarity) and non-weighted (unweighted UniFrac) algorithms and shown as principal coordinate analysis (PCoA). Spearman correlations and similarity percentage analysis (SIMPER) were performed to determine significant differences between bacterial communities among genotype and diet groups.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 June 2011; accepted 14 May 2012.

1. Khan, Y. & Bhutta, Z. A. Nutritional deficiencies in the developing world: current status and opportunities for intervention. *Pediatr. Clin. North Am.* **57,** 1409–1441 (2010).
2. Weisstaub, G. & Araya, M. Acute malnutrition in Latin America: the challenge of ending avoidable deaths. *J. Pediatr. Gastroenterol. Nutr.* **47,** S10–S14 (2008).
3. Imai, Y. *et al.* Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436,** 112–116 (2005).
4. Crackower, M. A. *et al.* Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature* **417,** 822–828 (2002).
5. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nature Med.* **11,** 875–879 (2005).
6. Donoghue, M. *et al.* A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. *Circ. Res.* **87,** E1–E9 (2000).
7. Tipnis, S. R. *et al.* A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase. *J. Biol. Chem.* **275,** 33238–33243 (2000).
8. Camargo, S. M. *et al.* Tissue-specific amino acid transporter partners ACE2 and collectrin differentially interact with Hartnup mutations. *Gastroenterology* **136,** 872–882 (2009).
9. Kowalczuk, S. *et al.* A protein complex in the brush-border membrane explains a Hartnup disorder allele. *FASEB J.* **22,** 2880–2887 (2008).
10. Kleta, R. *et al.* Mutations in SLC6A19, encoding B0AT1, cause Hartnup disorder. *Nature Genet.* **36,** 999–1002 (2004).
11. Segal, I. *et al.* Rectal manifestations of pellagra. *Int. J. Colorectal Dis.* **1,** 238–243 (1986).
12. Okayasu, I. *et al.* A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* **98,** 694–702 (1990).

13. Cooper, H. S., Murthy, S. N., Shah, R. S. & Sedergran, D. J. Clinicopathologic study of dextran sulfate sodium experimental murine colitis. *Lab. Invest.* **69,** 238–249 (1993).
14. Osterreicher, C. H. *et al.* Angiotensin-converting-enzyme 2 inhibits liver fibrosis in mice. *Hepatology* **50,** 929–938 (2009).
15. Vickers, C. *et al.* Hydrolysis of biological peptides by human angiotensin-converting enzyme-related carboxypeptidase. *J. Biol. Chem.* **277,** 14838–14843 (2002).
16. Swirski, F. K. *et al.* Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325,** 612–616 (2009).
17. Darby, W. J., McNutt, K. W. & Todhunter, E. N. Niacin. *Nutr. Rev.* **33,** 289–297 (1975).
18. Attia, E., Wolk, S., Cooper, T., Glasofer, D. & Walsh, B. T. Plasma tryptophan during weight restoration in patients with anorexia nervosa. *Biol. Psychiatry* **57,** 674–678 (2005).
19. Danilczyk, U. *et al.* Essential role for collectrin in renal amino acid transport. *Nature* **444,** 1088–1091 (2006).
20. Stratigos, J. D. & Katsambas, A. Pellagra: a still existing disease. *Br. J. Dermatol.* **96,** 99–106 (1977).
21. Nenci, A. *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* **446,** 557–561 (2007).
22. Chalaris, A. *et al.* Critical role of the disintegrin metalloprotease ADAM17 for intestinal inflammation and regeneration in mice. *J. Exp. Med.* **207,** 1617–1624 (2010).
23. Ghosh, H. S., McBurney, M. & Robbins, P. D. SIRT1 negatively regulates the mammalian target of rapamycin. *PLoS ONE* **5,** e9199 (2010).
24. Bröer, A. *et al.* Impaired nutrient signaling and body weight control in a Na+ neutral amino acid cotransporter (Slc6a19)-deficient mouse. *J. Biol. Chem.* **286,** 26638–26651 (2011).
25. Farkas, S. *et al.* Rapamycin decreases leukocyte migration *in vivo* and effectively reduces experimentally induced chronic colitis. *Int. J. Colorectal Dis.* **21,** 747–753 (2006).
26. Reinisch, W. *et al.* A multicenter, randomized, double-blind trial of everolimus versus azathioprine and placebo to maintain steroid-induced remission in patients with moderate-to-severe active Crohn's disease. *Am. J. Gastroenterol.* **103,** 2284–2292 (2008).
27. Mastroianni, J. R. & Ouellette, A. J. Alpha-defensins in enteric innate immunity: functional Paneth cell alpha-defensins in mouse colonic lumen. *J. Biol. Chem.* **284,** 27848–27856 (2009).
28. Salzman, N. H. *et al.* Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunol.* **11,** 76–83 (2010).
29. Weichhart, T. Mammalian target of rapamycin: a signaling kinase for every aspect of cellular life. *Methods Mol. Biol.* **821,** 1–14 (2012).
30. Abu-Qurshin, R. *et al.* Crohn's disease associated with pellagra and increased excretion of 5-hydroxyindolacetic acid. *Am. J. Med. Sci.* **313,** 111–113 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all members of our laboratories for discussions. T. Hashimoto was supported by grants from the European Respiratory Society, EuGeneHeart, and SENSHIN. T.P. was supported by a Marie Curie IIF. J.M.P. was supported by grants from IMBA, the Austrian Ministry of Sciences, the Austrian Academy of Sciences, GEN-AU (AustroMouse), an EU ERC Advanced Grant, and the EU network grants EuGeneHeart, ApoSys and INFLA-Care. P.R. was supported by the BMBF Network 'Systematic genomics of chronic inflammation', the DFG Cluster of Excellence Inflammation at Interfaces, SPP1399, SFB877, and the DFG project RO1394.

Author Contributions T. Hashimoto and T.P. performed most experiments, together with J.T., H.I., M.P., V.S., R.H. and T. Hanada; A. Rehman, P.R., S.L. and S.S. performed sequencing and analysis of the gut microbiome. S.M.R.C., D.S, F.V. and H.C. analysed collectrin and ACE2 expression in gut and kidney. A. Richter and B.W. performed amino acid analysis in serum. K.K. and A.F. provided essential mouse strains. J.M.P. and P.R. coordinated the project, and together with T. Hashimoto and T.P. wrote the manuscript and designed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.P. (josef.penninger@imba.oeaw.ac.at) or P.R. (p.rosenstiel@mucosa.de).

METHODS

Mice. *Ace2* (ref. 4), *Agtr1a* (ref. 31), *Apelin* (ref. 32), $Apj^{-/-}$ (ref. 33) and *Collectrin* (ref. 19) mutant mice have been described previously. All mutant strains were backcrossed onto a C57BL/6 background more than 6 generations. $Apj^{-/-}$ mice were on a mixed C57BL/6 × CBA background. In all experiments, only sex-, age- and background-matched mice were used as controls. Mice were handled and all experiments performed in accordance with institutional guidelines.

Colitis models. Acute colitis in C57BL/6 background mice was induced with 1.5% (w/v) DSS (molecular mass 36–50 kDa; MP Biologicals) dissolved in drinking water ad libitum for the experimental days 1–7 followed by normal drinking water until the end of the experiment. For mixed C57BL/6 × CBA background mice, colitis was induced by feeding mice 5% (w/v) DSS. The DSS solutions were made fresh every day. TNBS colitis was induced by intrarectal administration of 0.1 ml TNBS (5 mg per mouse in 40% ethanol), through a feeding needle (Fine Science Tools) 3.5 cm proximal to the anal verge as described previously[34]. Body weight, diarrhoea, and occult blood in stool (Hemoccult test) were determined daily. Diarrhoea scores were: 0, no diarrhoea; 1, mild diarrhoea; 2, severe watery diarrhoea; 3, mild diarrhoea with blood; 4, severe watery diarrhoea with blood. Colon length was determined at the end of the experiments. Crypt scores, inflammation scores, and disease activity indexes were determined as previously described[13]. For PFD experiment, normal chow (20.8% protein) and PFD (<0.2% protein) were obtained from Ssniff GmbH. Mice were fed the protein-free, isocaloric diet for 3 days before start of the 1% (w/v) DSS treatment. Owing to the severe diarrhoea in PFD fed mice, the experiment had to be stopped for ethical reasons on day 4. For nicotinamide (NAM) treatment, mice were given NAM (Sigma) in their drinking water ($0.4\,g\,l^{-1}$) from 3 days before the first DSS treatment until the end of the experiment. For rapamycin (RAPA) treatment, rapamycin ($2\,mg\,kg^{-1}\,d^{-1}$; LC Laboratories) or vehicle were administered intraperitoneally (i.p.) daily for 6 days with or without NAM treatment. The solvent for rapamycin was 0.2% sodium carboxymethylcellulose, 0.25% polysorbate-80 in water. For tryptophan rescue experiments, standard food was supplemented with glycyl-L-tryptophan (Gly-Trp) hydrate (G0144, TCI Europe) (10 mg per g dry food) daily for 10 days followed by 1.5% (w/v) DSS treatment. Tryptophan-free diet was purchased from Oriental Yeast and was fed for 18 days before DSS challenge. For antibiotic treatment, 4-week-old mice were placed on broad spectrum antibiotics ampicillin ($1\,g\,l^{-1}$) and neomycin ($0.5\,g\,l^{-1}$) in drinking water for 4 weeks followed by 1.5% (w/v) DSS treatment. For microbiota transplantation, caecal and ileum extracts were pooled from *Ace2* mutant mice and ACE2 expressing littermates, suspended in 3 ml PBS, and administered (0.1 ml per mouse) immediately to sterile-packed 4-week-old Swiss-Webster germ-free mice (Taconic). Transplanted mice were maintained in sterile cages for 2 weeks followed by 5% (w/v) DSS treatment. For ACE2 reconstitution, mice received intraperitoneal injections of recombinant mouse ACE2 protein ($0.1\,mg\,kg^{-1}$) (R&D) daily during 1.5% (w/v) DSS treatment.

Serum amino acid analyses. Heparinized serum was mixed with an equal part of a solution containing $20\,\mu g\,ml^{-1}$ each of norleucine (internal standard for Ala, Gly and Val), norvaline (for Leu, Ile and Pro) and parachloro-phenylalanine (for all other amino acids) and deproteinized by ultrafiltration through a 10 kDa cutoff regenerated cellulose ultrafiltration membrane (Millipore). Amino acids were then purified by cation exchange, converted to their chloroformate derivatives and separated and quantified by GC/MS in the selected ion monitoring (SIM) mode. Tryptophan ELISA kit (LDN, BA E-2700) was used to assess Trp serum levels upon Gly-Trp treatment.

Purification of intestinal epithelial cells and mucosal cells. After washing with PBS, segments of small intestines were placed in 1.5 mM EDTA in Hank's balanced salt solution (HBSS) without calcium and magnesium and tumbled for 10 min at 37 °C. The mucosa was incubated again with HBSS/EDTA for 10 min at 37 °C. The supernatant containing the crypts was collected. The number and viability of the crypt epithelial cells were determined by 0.1% trypan blue exclusion. Purity of the epithelial cell preparation was checked by haematoxylin and eosin staining. For mucosa preparations jejunum, ileum, or colon were inverted and mucosa was scraped off with a scalpel.

Quantitative RT–PCR. Total RNA was prepared from isolated small intestinal epithelial cells using the RNeasy Mini Kit (Qiagen) in accordance with the manufacturer's instructions. Total RNA (0.2 μg) was subjected to reverse transcription using random hexamers (Roche) and SuperscriptII (Invitrogen) followed by quantitative PCR analysis. The following primers were used: GAPDH, 5′-TCATCAACGGGAAGCCCATCAC-3′ (forward) and 5′-AGACTCCACGACATACTCAGCACCG-3′ (reverse); β-actin, 5′-GGCTGTATTCCCCTCCATCG-3′ (forward) and 5′-CCAGTTGGTAACAATGCCATGT-3′ (reverse); DEFA1, 5′-TCCTCCTCTCTGCCCTYGTCCTG-3′ (forward) and 5′-CTCTTCTCCTGGCTGCTCCTC-3′ (reverse); DEFA3, 5′-GCTAGGGAGCACTTGTTTGC-3′ (forward) and 5′-TTGTTTGAGGAAAGGAGGCA-3′ (reverse); DEFA4, 5′-CCAGGGGAAGATGACCAGGCTG-3′ (forward) and 5′-TGCAGCGACGATTTCTACAAAGGC-3′ (reverse); DEFA5, 5′-AGGCTGATCCTATCCACAAAACAG-3′ (forward) and 5′-TGAAGAGCAGACCCTTCTTGGC-3′ (reverse); DEFA-RS1, 5′-CACCACCCAAGCTCCAAATACACAG-3′ (forward) and 5′-ATCGTGAGGACCAAAAGCAAATGG-3′ (reverse); MMP7, 5′-ACCCTGTTCTGCTTTGTGTGTC-3′ (forward) and 5′-TCTGAGCCTGTTCCCACTGATG-3′ (reverse); Angiogenin1 (ANG1), 5′-CCATGACGCCAAGCCAAAGG-3′ (forward) and 5′-GCTTCCATTCGCTCCACAGATG-3′ (reverse); Angiogenin4 (ANG4), 5′-GCTGGGTCTGGTTGTGATTCC-3′ (forward) and 5′-AGGCGAGGTTAGCTTTCTTTCC-3′ (reverse); FABP2, 5′-AGCTCGGTGTAAACTTTCCCTAC-3′ (forward) and 5′-AAACCTCTCGGACAGCAATCAG-3′ (reverse); LYZ1, 5′-GCCAAGGTCTACAATCGTTGTGAGTTG-3′ (forward) and 5′-CAGTCAGCCAGCTTGACACCACG-3′ (reverse); LYZ2, 5′-GGCTGGCTACTATGGAGTCAGCCTG-3′ (forward) and 5′-GCATTCACAGCTCTTGGGGTTTTG-3′ (reverse); PLA2G2E, 5′-AGGATTCCCCCAAGGATGCCAC-3′ (forward) and 5′-CAGCCGTTTCTGACAGGAGTTCTGG -3′ (reverse); ACE, 5′-TCATCATCCAGTTCCAGTTCCA-3′ (forward) and 5′-CGGTGACGAGCCATTCTGT-3′ (reverse); ACE2, 5′-TGGTCTTCTGCCATCCGATT-3′ (forward) and 5′-CCATCCACCTCCACTTCTCTAA-3′ (reverse); B^0AT1, 5′-AACGCTCATGTATAGCATCTGG-3′ (forward) and 5′-CAGCCACAGTGACCACAAC-3′ (reverse); TLR1, 5′-AAGAACTCAGGCGAGCAGAG-3′ (forward) and 5′-TCTGGTAAGGTTTGCGTTTGG-3′ (reverse); TLR2, 5′-CACTTCCAGGTCTTCAGTCTTC-3′ (forward) and 5′-ACTATGATTGCGGACACATCTC-3′ (reverse); TLR3, 5′-TTGCGTTGCGAAGTGAAGAA-3′ (forward) and 5′-AGACAGATACCTCAGGCTTGG-3′ (reverse); TLR4, 5′-CAGAACTTCAGTGGCTGGATT-3′ (forward) and 5′-TGTACTAGGTTCGTCAGATTGG-3′ (reverse); TLR5, 5′-GAGTGAGGTCAGTCCTGGAG-3′ (forward) and 5′-GAGATGAGGCGTCTGGAGAG-3′ (reverse); TLR6, 5′-TACCGTCAGTGCTGGAAATAGA-3′ (forward) and 5′-TTCTGTCTTGGCTCATGTTGC-3′ (reverse); TLR7, 5′-AGAACCACTCTATGCCTTCAAG-3′ (forward) and 5′-GTCTCTTCCGTGTCCACATC-3′ (reverse); TLR8, 5′-CCTTCCTTTGTCTATAGAACATGGA-3′ (forward) and 5′-CACAAGGATAGCTTCTGGAATAGT-3′ (reverse); TLR9, 5′-AGAATCCTCCATCTCCCAACAT-3′ (forward) and 5′-CAGACTTCAGGAACAGCCAAT-3′ (reverse).

Western blotting and ELISA. Western blotting was performed with standard protocols using isolated primary small intestinal epithelial cells. Primary antibodies reactive to phosphorylated S6 (2215; Cell Signaling), S6 (2217; Cell Signaling), and β-actin (Sigma) were used. Blots were washed three times in TBST for 30 min, incubated with horseradish peroxidase-conjugated secondary antibodies (1:5,000 dilution; Jackson ImmunoResearch) for 20 min, washed three times in TBST, and visualized with enhanced chemiluminescence. Small intestine was collected for ELISA assays to detect phosphorylated p70S6K or total p70S6K (DYC896, DYC8962; R&D) and colon was collected to determine angiotensin II levels (EK-002-12; Phoenix). To analyse membrane proteins, total membranes were prepared using mannitol buffer (200 mM mannitol, 80 mM HEPES, 41 mM KOH, pH 7.5 in the presence of protease inhibitor cocktail, Sigma). After homogenization of the samples using MagNA Lyser green beads (Roche), total membranes were attained by dual consecutive centrifugation steps (100*g* and 100,000*g*) at 4 °C. 50 μg of total membrane were used per lane for western blotting. The following primary antibodies were used: anti-Ace2 and anti-Ace (R&D Systems, 1:1,000), anti-B^0AT1 (Pineda, 1:1,000), anti-mouse β-actin (Sigma, 1:10,000). Antibody binding was detected with Immobilon Western Chemiluminescent HRP or AP substrates (Millipore) and chemiluminescence visualized with a DIANA III camera (Raytest, Dietikon).

Histology and immunohistochemistry. For histological analysis, 2–4-μm-thick sections were stained with haematoxylin and eosin. Immunohistochemistry was performed using the automated Ventana system. Sections were incubated with antibodies against phosphorylated S6 (Cell Signaling), Ace2 (R&D), Ki67 (Novacastra), and cleaved caspase 3 (Cell Signaling). For phosphorylated S6 and cleaved caspase 3 staining, a multimer based detection system (UltraMap) was used. In addition, immunofluorescence was performed on cryosections of the kidney and small intestine using anti-Ace2 antibodies at 1:100, anti-B^0AT1 at 1:200, and anti-collectrin at 1:1,000 followed by Alexa488-conjugated anti-goat IgG or Alexa594-labelled anti-rabbit IgG.

DNA extraction, pyrosequencing, sequence analysis. Genomic DNA from terminal ileum was extracted by MO BIO PowerSoil DNA Isolation kit with modifications: solution C1 and 20 μl of Proteinase K were added to faeces and incubated for 2 h at 50°C to enhance the lysis. Extracted DNA was quantified by using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Variable region V1-V2 of the 16S rRNA gene was amplified using the forward primer 5′-**CTATGCGCCTTGCCAGCCCGC**tcagTC<u>AGAGTTTGATCCTGGCTCAG</u>-3′. Sequence in bold is 454 Life Sciences primer B, the underlined sequence represents the broadly conserved bacterial primer 27F. A two-base linker sequence ('TC')

and a four base key (lower case) were added. The reverse primer (5′- **CGTATCGCCTCCCTCGCGCCA**tcagXXXXXXXXXXCATGCTGCCTCCCGTAGGAGT-3′) contained the 454 Life Sciences primer A sequence (bold) a unique 10 base multiplex identifier (MIDs designated as XXXXXXXXXX) to tag each PCR product, the broad-range bacterial primer 338R, and a 'CA' linker sequence inserted between the MIDs and the rRNA gene primer. In addition, a four base key (lower case) was added before the MIDs as described earlier[35]. Replicate PCR reactions were performed for each sample and the respective negative controls. Each reaction consisted of 1 μl (10 pmol μl^{-1}) of each primer, 40 ng of DNA, 1 μl of dNTPs mix (10 mM stock), 1.0 units of Phusion Hot Start II High-Fidelity DNA Polymerase, and 10 μl of 5× Phusion HF buffer. The final volume of reaction was adjusted to 50 μl. Amplification was performed using an initial denaturation of 3 min at 98 °C followed by 30 cycles, denaturation at 98 °C for 10 s, annealing at 55 °C for 30 s and elongation at 72 °C for 30 s. Final extension was at 72 °C for 10 min. Amplified products were run on an agarose gel, specific bands excised and amplicons purified using the Qiagen gel purification kit (Qiagen). Concentrations of the eluted and purified amplicons were measured using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Equal amounts of PCR products were mixed in a single tube and sequenced using Roche 454 Titanium chemistry. To generate negative control templates for the PCR reaction, water was taken as a sample during DNA extraction; in all cases, the PCR reactions from these templates were negative.

Sequence analysis. After pyrosequencing, all sequence reads were screened and filtered for quality and length using PANGEA[36]. Sequences were trimmed and binned by samples using the specific MIDs as barcode. Sequences not meeting the following quality criteria were excluded from further analysis: no perfect match with MID sequences or specific primers; reads shorter than 200 bases; mean quality score ≥25; the presence of any ambiguous base; and/or more than 8 homopolymers. Remaining sequences were aligned using the curated SILVA seed database (bacterial reference alignment) as a template with the Needleman-Wunsch algorithm[37]. Sequences not matching with defined core regions and putative chimaeric sequences were detected by using the uchime[38] command implemented in Mothur[39] and were subsequently removed. Sequences were further confirmed to originate from bacteria using the Ribosomal Data base project (RDP) classifier with 60% bootstrap threshold[40]. For all the downstream analysis, 1,000 sequences per samples were randomly sampled to normalize the number of reads distribution. In order to reduce the overestimation of distances between sequences, leading and trailing periods from aligned files were removed. Pairwise distances between sequences were calculated, gaps of any length were treated as single event or mismatch. Sequences with ≥97% similarity were clustered using average neighbour algorithm in to species level operational taxonomical units (OTUs)[41]. Phylogenetic affiliation of each OTU was performed by using RDP taxonomy and a template file. To visualize the microbial composition and structure in related on host genotype or treatment, distance matrices between samples were generated based on weighted (Bray–Curtis similarity) and non-weighted (unweighted UniFrac) algorithms and visualized using principal coordinate analysis (PCoA). OTUs contributing for the shifting of the samples along the two axes were ascertained by calculating the Spearman correlation of each OTU with the two axes in the PCoA[39]. Similarity percentage analysis (SIMPER)[42] was performed to determine the driving OTUs imparting significant differences among bacterial community composition among genotype and diet groups. The analysis was performed on Bray–Curtis similarity coefficient with the program PAST (Paleontological Statistics, ver. 2.14)[43].

Bone marrow transplantation. Eight week old *Ace2* mutant and *Ace2* wild-type mice were lethally irradiated with 980 cGy (Gammacell 40). Bone marrow cells were harvested from 8–9-week-old donor *Ace2* mutant or *Ace2* wild-type mice by flushing the femurs and tibias with DMEM. Recipient mice received 5×10^6 bone marrow cells in 250 μl DMEM from male donor mice by tail vein injection 6 h after irradiation. Six weeks after transplantation, the genotypes of haematopoietic cells were determined by PCR using peripheral blood. The PCR primer set for the mutant *Ace2* allele was as follows: 5′-CCAGCTCATTCCTCCCACTC-3′ and 5′-CCGGCTGCTCTTTGAGAGGACA-3′. The PCR primer set for wild-type *Ace2* was: 5′-CTTCATTGGCTCCGTTTCTTAGC-3′ and 5′-CCGGCTGCTCTTTGAGAGGACA-3′.

Calorimetry cages. Average food intake was measured using indirect calorimetry cages from TSE Systems over a period of seven days following 48 h of acclimatization. Data were analysed with proprietary Phenomaster Analysis Software (TSE Systems) and Microsoft Excel.

Statistical analyses. All data are shown as mean ± s.e.m. Measurements at single time points were analysed by ANOVA and if significant, further analysed by a two-tailed *t*-test. $P < 0.05$ was considered to indicate statistical significance.

31. Ito, M. *et al.* Regulation of blood pressure by the type 1A angiotensin II receptor gene. *Proc. Natl Acad. Sci. USA* **92,** 3521–3525 (1995).
32. Kuba, K. *et al.* Impaired heart contractility in Apelin gene-deficient mice associated with aging and pressure overload. *Circ. Res.* **101,** e32–e42 (2007).
33. Ishida, J. *et al.* Regulatory roles for APJ, a seven-transmembrane receptor related to angiotensin-type 1 receptor in blood pressure in vivo. *J. Biol. Chem.* **279,** 26274–26279 (2004).
34. McCafferty, D. M., Miampamba, M., Sihota, E., Sharkey, K. A. & Kubes, P. Role of inducible nitric oxide synthase in trinitrobenzene sulphonic acid induced colitis in mice. *Gut* **45,** 864–873 (1999).
35. Rehman, A. *et al.* Nod2 is essential for temporal development of intestinal microbial communities. *Gut* **60,** 1354–1362 (2011).
36. Giongo, A. *et al.* PANGEA: pipeline for analysis of next generation amplicons. *ISME J.* **4,** 852–861 (2010).
37. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
38. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
39. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
40. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
41. Schloss, P. D. & Handelsman, J. Status of the microbial census. *Microbiol. Mol. Biol. Rev.* **68,** 686–691 (2004).
42. Clarke, K. R. Nonparametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18,** 117–143 (1993).
43. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. Past: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**, art. 4 (2004).

LETTER

doi:10.1038/nature11286

Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy

N. M. Archin[1], A. L. Liberty[1], A. D. Kashuba[1], S. K. Choudhary[1], J. D. Kuruc[1], A. M. Crooks[1], D. C. Parker[1], E. M. Anderson[2], M. F. Kearney[2], M. C. Strain[3], D. D. Richman[3], M. G. Hudgens[1], R. J. Bosch[4], J. M. Coffin[2], J. J. Eron[1], D. J. Hazuda[5] & D. M. Margolis[1]

Despite antiretroviral therapy, proviral latency of human immunodeficiency virus type 1 (HIV-1) remains a principal obstacle to curing the infection[1]. Inducing the expression of latent genomes within resting CD4$^+$ T cells is the primary strategy to clear this reservoir[2,3]. Although histone deacetylase inhibitors such as suberoylanilide hydroxamic acid (also known as vorinostat, VOR) can disrupt HIV-1 latency *in vitro*[4–6], the utility of this approach has never been directly proven in a translational clinical study of HIV-infected patients. Here we isolated the circulating resting CD4$^+$ T cells of patients in whom viraemia was fully suppressed by antiretroviral therapy, and directly studied the effect of VOR on this latent reservoir. In each of eight patients, a single dose of VOR increased both biomarkers of cellular acetylation, and simultaneously induced an increase in HIV RNA expression in resting CD4$^+$ cells (mean increase, 4.8-fold). This demonstrates that a molecular mechanism known to enforce HIV latency can be therapeutically targeted in humans, provides proof-of-concept for histone deacetylase inhibitors as a therapeutic class, and defines a precise approach to test novel strategies to attack and eradicate latent HIV infection directly.

Among the many important aims of future HIV research is the development of therapies of finite duration capable of eradicating HIV infection. The persistence of quiescent HIV infection within a small population of long-lived CD4$^+$ T cells is currently a major obstacle to this goal[1]. Histone deacetylases (HDACs) are recruited to the HIV long terminal repeat (LTR) promoter, establishing one of several restrictions that can limit LTR expression and maintain viral latency[2,3]. Deacetylated LTR chromatin seems to play a key contributory role in regulating HIV expression, and especially in maintaining proviral quiescence and latency. *In vitro*, HDAC inhibitors have been shown to disrupt latent proviral HIV infection in both cell culture models and *ex vivo* assays using cells from HIV-1-infected patients. Although disrupting latency has been proposed as part of a strategy to eradicate HIV infection, previous studies using the weak HDAC inhibitor valproic acid did not consistently demonstrate a marked depletion of resting cell infection[7–11] in patients on antiretroviral therapy (ART). However, the effects measured in these studies are significantly downstream of the molecular site of action of HDAC inhibitors, and thus the proximal pharmacodynamic measures of HDAC inhibitor activity and HIV-1 expression were not evaluated. Here we show that HDAC inhibitors disrupt the latency of proviral genomes within resting CD4$^+$ T cells, establishing the first (to our knowledge) class of drugs that could lead to the eradication of HIV infection.

VOR is a potent HDAC inhibitor used to treat human malignancies. At clinically relevant concentrations, VOR inhibits the class I HDACs most important for repression of HIV expression[4,12]; it also induces LTR expression and virus production *in vitro* from the resting CD4$^+$ T cells of HIV-positive patients on ART with levels of plasma HIV RNA below the detection limit (BDL)[5,6,13]. As the most proximal measure of effect on latent infection is expression of HIV-1 RNA, we developed a sensitive assay to enable a direct measurement of unspliced *gag* HIV RNA within the resting CD4$^+$ T cells of HIV-infected patients. The assay has a limit of detection of 1 copy per million resting CD4$^+$ T cells, and a limit of quantification of 10 copies per million resting CD4$^+$ T cells.

To evaluate the effect of VOR on latent infection *in vivo*, HIV-infected patients receiving stable ART with plasma HIV-1 RNA <50 copies per ml for at least 6 months and a CD4 count >300 μl^{-1} were enrolled following informed consent. To demonstrate that it was ethical to expose patients to an experimental agent with potential risk in a study with no proven clinical benefit for the individual, we validated the ability of this assay of HIV RNA within resting CD4$^+$ T cells to measure HIV expression at baseline, and to detect up-regulation of HIV expression in resting cells from each patient after physiological exposure to VOR.

Patients maintained suppressive ART, and purified populations of resting CD4$^+$ T cells were obtained by continuous-flow leukapheresis and negative selection in an immunomagnetic column[7]. To establish a baseline, we measured the mean quantity of HIV-1 *gag* RNA in pools of 1 million resting CD4$^+$ T cells immediately after their isolation from patients. To measure validated biomarkers of VOR effect in peripheral blood mononuclear cells (PBMCs) of patients, we performed parallel assays of total cellular histone acetylation and measured histone acetylation by chromatin immunoprecipitation (ChIP) at the human p21 gene promoter, a gene known to upregulate chromatin acetylation after VOR exposure[14]. Then to model the effect of a clinical dose of VOR, multiple replicate pools of 1 million resting CD4$^+$ T cells were incubated in complete media alone, with 335 nM VOR, or with 3 μg phytohaemagglutinin (PHA) and 60 U interleukin-2 (IL-2) for 6 h. VOR conditions were selected to mimic the unbound drug exposure expected after a single 400 mg dose of VOR *in vivo*[5].

Validation assays were performed in resting CD4$^+$ T cells isolated by leukapheresis from 16 patients with plasma HIV RNA BDL (Fig. 1a). In each patient a total of 48–72 million highly purified resting CD4$^+$ T cells were studied; that is, 12–48 million cells in each condition, depending on cell availability. In 9 patients following 6 h of culture of 16–24 million cells without stimulation in media alone, HIV *gag* RNA was quantifiable at a mean level of 52 ± 32 copies per million cells. However, in the other 7 patients in whom 12–24 million cells were studied (Fig. 1a), HIV RNA was not quantifiable at a limit of 10 copies per million cells, although in all but 2 of these patients RNA was detected but not quantifiable (>0 but <10 copies per million cells).

Following *in vitro* exposure to 335 nM VOR for 6 h, HIV RNA expression was significantly upregulated in 8 of 9 patients in whom resting CD4$^+$ T cell HIV RNA was quantifiable without HDAC inhibitor exposure, and also in 3 of 7 patients in whom cell-associated HIV RNA was <10 copies per million cells before HDAC inhibitor exposure. In

[1]The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [2]HIV Drug Resistance Program, NCI, NIH, Frederick, Maryland 21702, USA. [3]VA San Diego Healthcare System and University of California San Diego, San Diego, California 92093, USA. [4]Harvard School of Public Health, Boston, Massachusetts 02115, USA. [5]Merck Research Laboratories, White Horse Junction, Pennsylvania 08889, USA.

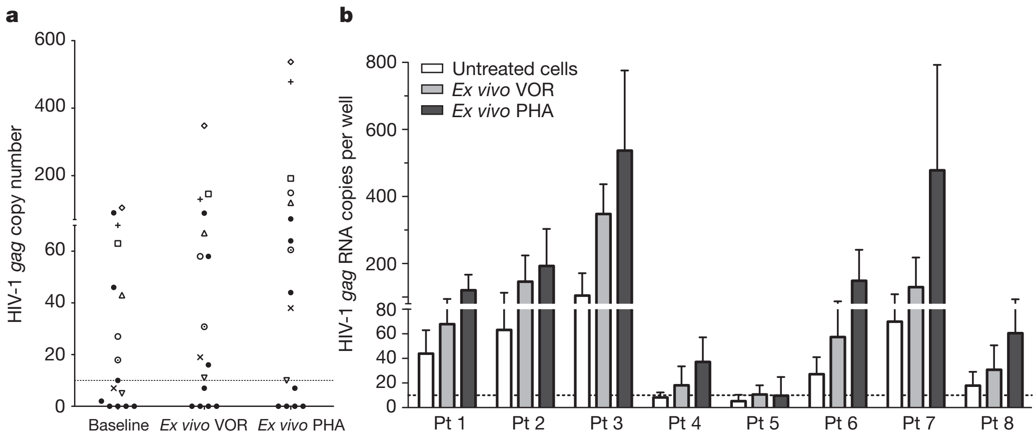


Figure 1 | The relative HIV-1 RNA copy number in resting CD4$^+$ T cells of 16 ART-treated HIV-positive patients with plasma HIV RNA BDL: Values are calculated by cycle number, and the limit of quantification of cell-associated RNA is 10 copies. Cells were cultured alone (untreated), with VOR 335 nM ('*ex vivo* VOR'), or activated with 3 μg ml^{-1} PHA and 60 U ml^{-1} IL-2 for 6 h ('*ex vivo* PHA'). **a**, Data from 16 patients; **b**, data from eight of those patients (Pt 1–8) who later received *in vivo* dosing (mean and s.d.). The patients in **b** are identified by the following symbols in **a**: patient 1, upright open triangle; 2, open square; 3, open diamond; 4, cross; 5, inverted open triangle; 6, open circle; 7, 'plus' sign; 8, open dotted circle.

all 11 patients in whom expression was induced by VOR, HIV RNA expression was also induced after 6 h of exposure to PHA. Levels of HIV RNA induction were similar after 6 h of exposure to PHA or VOR, although in other experiments (data not shown) levels of HIV RNA plateaued after 6 h of exposure to VOR, but continued to increase when PHA exposure was extended for up to 16 h.

Of the 11 eligible patients in whom an induction of HIV RNA expression following *in vitro* exposure to VOR could be measured, eight patients (Fig. 1b; CD4 count 432–1,147 μl^{-1}, mean 713 μl^{-1}) continued their long-term ART (tenofovir, emtricitabine and efavirenz in all) and agreed to receive a single dose of VOR at 200 mg to ascertain tolerability. Global cellular histone acetylation in PBMCs was unchanged up to 24 h after dosing, and histone acetylation at the human p21 gene promoter was modestly increased in only two patients (data not shown).

Two to four weeks later the safety and tolerability of a 400 mg dose of VOR was assessed, including acetylation measurements, and VOR plasma concentrations measured serially up to 24 h after dosing (Fig. 2). Maximum VOR concentrations of 244 ng ml^{-1} (median, range 153–301) occurred a median of two hours (range 0.5–4) after dosing. VOR pharmacokinetics were similar to those reported in other populations[15,16]. This exposure resulted in a significant increase ($P < 0.01$) in acetylation of total cellular histone H3 (median 1.6-fold) in all eight patients, and trend towards increased acetylation of histones at the human p21 gene as measured by ChIP in patients 2, 3, 4, 5, and 7 for whom sufficient cells were available (Fig. 2).

Four to five weeks later, a second dose of 400 mg of VOR was administered and resting CD4$^+$ T cells collected 4–7 h later, after the previously measured peak VOR level. We measured HIV RNA in multiple pools of resting CD4$^+$ T cells within this window of time, when cellular biomarkers of increased acetylation had been previously documented, and induction of HIV LTR expression mediated by HDAC inhibitor activity would be possible. During exposure to VOR, an increase of 1.5- to 10.0-fold (mean 4.8) in expression of unspliced HIV-1 *gag* RNA within resting CD4$^+$ T cells was measured in all eight patients (Fig. 3). Expression following the 400 mg dose was significantly increased ($P < 0.01$) when compared to baseline levels of RNA expression for all patients. Patients 1 and 2 underwent leukapheresis 3 to 4 months after dosing and protocol completion to provide cells for other research, and in both cases a statistically

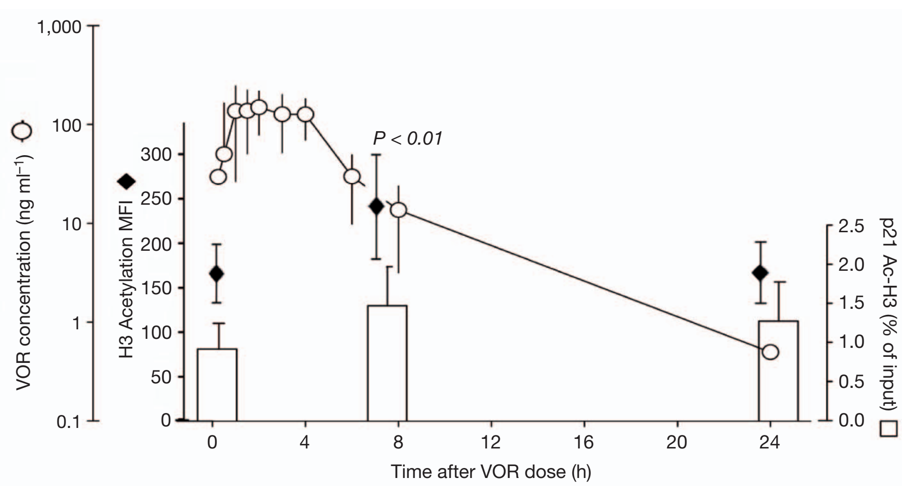


Figure 2 | VOR exposure and histone acetylation. Median VOR plasma concentrations in patients 1–8 after a single 400 mg oral dose (open circles; error bars show range) are shown in comparison with mean fluorescence intensity (MFI) of total cellular acetylated histone H3 in PBMCs (filled diamonds; error bars show ±1 s.d.), and relative levels of acetylated histone H3 at the human p21 gene promoter in resting CD4$^+$ T cells (histogram; data show mean +1 s.d.). A significant increase ($P < 0.01$) in cellular acetylated histone H3 is seen at 8 h. A trend towards increased acetylation at the p21 gene is seen in the patients (2–5, 7) in whom sufficient cells were available for analysis.

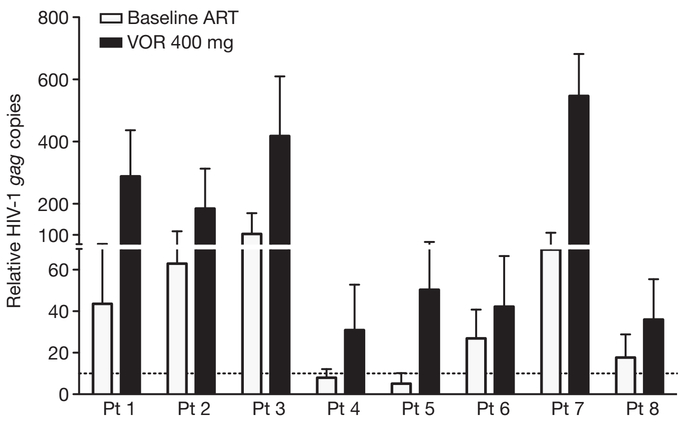


Figure 3 | VOR upregulates HIV RNA expression. The relative HIV-1 RNA copy number (mean +1 s.d.) measured in the resting $CD4^+$ T cells of eight HIV-positive patients with plasma HIV RNA BDL is shown on background ART and on ART following a single 400 mg oral dose of VOR. For each subject, the differences are significant ($P < 0.01$).

significant decline in HIV RNA expression per million resting $CD4^+$ T cells after VOR dosing was observed (patient 1: after VOR 290 copies, off-protocol 70 copies per million cells, $P < 0.001$; patient 2: after VOR 186 copies, off-protocol 105 copies per million cells, $P = 0.003$). These observations support the conclusion that the increase in HIV RNA expression was causally related to VOR exposure. This conclusion assumes that the isolation of RNA from resting $CD4^+$ T cells largely excludes actively infected cells, and reflects RNA expression in cells that are functionally defined as latently infected[17].

Throughout this limited exposure, VOR was well tolerated, with no adverse events greater than grade I; none of these effects were attributable to VOR. Additionally, we measured single-copy assay viraemia[18] on occasions before VOR exposure, and at 8 and 24 h after doses of 200, 400 and 400 mg of VOR (10 assays in total). Despite the upregulation of HIV RNA expression measured in circulating resting $CD4^+$ T cells, no significant change of low-level viraemia was observed. Median plasma HIV RNA (in copies per ml) by single-copy assay in these 8 patients was 13 (range 3 to 23), <1.0 (range <1 to 3), 2 (range <1 to 3), <1 (range <1 to 1), 3 (range 1 to 6), <1 (range <1 to 5), <1 (range <1 to 1.2), and <1 in all assays, with no consistent trend towards an increase in plasma HIV RNA after VOR exposure.

This study demonstrates that the quiescence of latent, integrated HIV provirus within resting $CD4^+$ T cells, a significant barrier to the eradication of HIV infection, can be disrupted by an achievable and tolerable exposure to an HDAC inhibitor. A single, clinically tolerable dose of VOR induces the expected biological effect—histone acetylation—consistent with HDAC inhibitor exposure within the PBMCs of HIV-infected, ART-treated patients. These effects are temporally associated with increased levels of HIV RNA expression detected within resting $CD4^+$ T cells, demonstrating that, at least for a period of time in some infected cells, all of the restrictions that limit the expression of latent proviral genomes have been overcome. Nevertheless, although histone deacetylation is associated with HIV RNA expression, the precise molecular mechanisms through which VOR mediates this effect remain to be fully explained.

HIV RNA induction *in vivo* often appeared to be of greater magnitude than that seen *in vitro*, perhaps due to an underestimation in the modelling of physiological exposure *in vitro*, or other phenomena induced by drug exposure *in vivo*. Further testing will be required to determine if the *in vitro* assay presented here is predictive of a subset of patients who do not respond to VOR, or if such pre-screening is unnecessary. Nevertheless, assays of HIV RNA associated with resting $CD4^+$ T cells may be useful in the pre-clinical and clinical testing of new and combination approaches to disrupt HIV latency.

Precise studies will be required to determine which dosing regimens of VOR or other HDAC inhibitors are safe and result in sustained disruption of HIV latency. We did not observe an alteration of low-level viraemia. This may be due to the very low levels of plasma viraemia present in these patients at baseline, that our sampling missed the brief effect of a single exposure to VOR, or that stable, low-level viraemia reflects virion production from sources other than resting $CD4^+$ T cells. Further, a limited evaluation did not reveal a substantial reduction in the frequency of replication-competent HIV within resting $CD4^+$ T cells (data not shown). This is not surprising, as such an effect is likely to require more than a single dose of VOR, or additional interventions to clear infected cells. These are important goals for future studies.

VOR, like the HDAC inhibitor in wide clinical use, valproic acid, is deemed a mutagen as predicted by the Ames test in bacteria, although DNA damage induced by VOR is known to be repaired in normal human cells[19,20]. Mutations in normal cells are not observed following long-term growth in the presence of VOR (P. A. Marks, personal communication). Nevertheless, the risks and benefits of attempts to eradicate HIV infection will have to be carefully weighed. Whereas it remains to be seen if the use of VOR in combination with suppressive ART is sufficient to result in the depletion of latently infected resting $CD4^+$ T cells, or whether additional interventions are required, these findings demonstrate that therapy targeted at persistent, latent infection within resting $CD4^+$ T cells is feasible, and open the way for the development of HDAC inhibitors with improved specificity, potency and safety profiles for the selective targeting of latent proviral genomes.

METHODS SUMMARY

HIV-infected patients receiving stable, standard-of-care ART with plasma HIV-1 RNA <50 copies per ml and a CD4 count of $>300\,\mu l^{-1}$ for at least 6 months were enrolled following informed consent. Studies were approved by the UNC institutional biomedical review board and the Food and Drug Administration.

Leukapheresis was performed and resting $CD4^+$ T cells isolated[8]. *Ex vivo* HIV-1 RNA production in resting $CD4^+$ T cells was measured following incubation in media containing 335 nM VOR, or $3\,\mu g\,ml^{-1}$ PHA and $60\,U\,ml^{-1}$ IL-2, or media alone for 6 h. To provide a baseline for the *in vivo* response to VOR, other resting $CD4^+$ T cells were immediately frozen and stored.

Total RNA was isolated from 12 to 48 pools of 1 million resting cells, duplicate pools of cDNA were synthesized, and duplicate PCR amplification performed[21]. A standard curve was generated for each PCR reaction from *in vitro* transcribed RNA[22]. Results of the four PCR replicates representing each of the original 12 to 48 pools of RNA were averaged and the standard deviation determined for each condition. The Wilcoxon rank sum test was used to calculate the statistical significance of all comparisons between conditions.

To measure total histone acetylation, PBMCs were fixed, permeabilized and stained with anti-acetyl histone H3 and a FITC-conjugated secondary antibody, and analysed by flow cytometry. ChIP was performed as previously[4] but 1×10^6 sonicated PBMCs were used for immunoprecipitation with anti-acetyl-histone H3 or pre-immune globulin, and PCR of performed using primers targeting the human p21 promoter[23].

Pharmacokinetic assays of VOR were performed before dosing and at 0.25, 0.50, 1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0 and 24 h after dosing. VOR was measured using a modification of the method of ref. 15.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 February; accepted 7 June 2012.

1. Blankson, J. N., Persaud, D. & Siliciano, R. F. The challenge of viral reservoirs in HIV-1 infection. *Annu. Rev. Med.* **53,** 557–593 (2002).
2. Margolis, D. M. Histone deacetylase inhibitors and HIV latency. *Curr. Opin. HIV AIDS* **6,** 25–29 (2011).
3. Van. Lint, C. *et al.* Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J.* **15,** 1112–1120 (1996).
4. Archin, N. M. *et al.* Expression of latent human immunodeficiency type-1 is induced by novel and selective histone deacetylase inhibitors. *AIDS* **23,** 1799–1806 (2009).
5. Archin, N. M. *et al.* Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. *AIDS Res. Hum. Retroviruses* **25,** 207–212 (2009).

6. Contreras, X. *et al.* Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells. *J. Biol. Chem.* **284,** 6782–6789 (2009).
7. Lehrman, G. *et al.* Depletion of latent HIV infection *in vivo*: a proof of concept study. *Lancet* **366,** 549–555 (2005).
8. Archin, N. M. *et al.* Valproic acid without intensified antiviral therapy has limited impact on persistent HIV infection of resting CD4+ T cells. *AIDS* **22,** 1131–1135 (2008).
9. Siliciano, J. D. *et al.* Stability of the latent reservoir for HIV-1 in patients receiving valproic acid. *J. Infect. Dis.* **195,** 833–836 (2007).
10. Sagot-Lerolle, N. *et al.* Prolonged valproic acid treatment does not reduce the size of latent HIV reservoir. *AIDS* **22,** 1125–1129 (2008).
11. Archin, N. M. *et al.* Antiretroviral intensification and valproic acid lack sustained effect on residual HIV-1 viremia or resting CD4+ cell infection. *PLoS ONE* **5,** e9390 (2010).
12. Keedy, K. S. *et al.* A limited group of class I histone deacetylases act to repress human immunodeficiency virus type-1 expression. *J. Virol.* **83,** 4749–4756 (2009).
13. Edelstein, L. C., Micheva-Viteva, S., Phelan, B. D. & Dougherty, J. P. Activation of latent HIV type 1 gene expression by suberoylanilide hydroxamic acid (SAHA), an HDAC inhibitor approved for use to treat cutaneous T cell lymphoma. *AIDS Res. Hum. Retroviruses* **25,** 883–887 (2009).
14. Gui, C. Y. *et al.* Histone deacetylase (HDAC) inhibitor activation of p21WAF1 involves changes in promoter-associated proteins, including HDAC1. *Proc. Natl Acad. Sci. USA* **101,** 1241–1246 (2004).
15. Du, L., Musson, D. G. & Wang, A. Q. High turbulence liquid chromatography online extraction and tandem mass spectrometry for the simultaneous determination of suberoylanilide hydroxamic acid and its two metabolites in human serum. *Rapid Commun. Mass Spectrom.* **19,** 1779–1787 (2005).
16. Rubin, E. H. *et al.* A study to determine the effects of food and multiple dosing on the pharmacokinetics of vorinostat given orally to patients with advanced cancer. *Clin. Cancer Res.* **12,** 7039–7045 (2006).
17. Hermankova, M. *et al.* Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes *in vivo. J. Virol.* **77,** 7383–7392 (2003).
18. Palmer, S. *et al.* New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **41,** 4531–4536 (2003).
19. Lee, J. H., Choy, M. L., Ngo, L., Foster, S. S. & Marks, P. A. Histone deacetylase inhibitor induces DNA damage, which normal but not transformed cells can repair. *Proc. Natl Acad. Sci. USA* **107,** 14639–14644 (2010).
20. Lee, J. H., Choy, M. L., Ngo, L., Venta-Perez, G. & Marks, P. A. Role of checkpoint kinase 1 (Chk1) in the mechanisms of resistance to histone deacetylase inhibitors. *Proc. Natl Acad. Sci. USA* **108,** 19629–19634 (2011).
21. Israel-Ballard, K, *et al.* TaqMan RT-PCR and VERSANT HIV-1 RNA 3.0 (bDNA) assay quantification of HIV-1 RNA viral load in breast milk. *J. Clin. Virol.* **34,** 253–256 (2005).
22. Robinson, L. H., Gale, C. V. & Kleim, J. P. Inclusion of full-length human immunodeficiency virus type 1 (HIV-1) gag sequences in viral recombinants applied to drug susceptibility phenotyping. *J. Virol. Methods* **104,** 147–160 (2002).
23. Radonić, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* **313,** 856–862 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by National Institutes of Health grants AI084553, AI095052 and AI096113, and by a grant from Merck & Co to D.M.M., National Institutes of Health grants RR024383 to the UNC TRaCS Institute, AI50410 to the UNC Center for AIDS Research, and an equipment grant from the James B. Pendleton Charitable Trust. Merck & Co provided VOR. The ClinicalTrials.gov identifier is NCT01319383. We thank N. P. Dahl, R. Sackmann, M. Cheema, A. Wiegand and N. White for technical support, Y. Park and the staff the UNC Blood Bank, R. J. Bedimo and C. M. van der Horst for study oversight, and J. Scepanski and A. Sugarbaker for patient follow-up and study coordination. We thank M. S. Cohen, J. V. Garcia-Martinez, W. C. Greene, J. Karn and R. F. Siliciano for discussions. Finally, we are grateful for the contributions of the patients who have participated in these studies.

Author Contributions N.M.A., A.L.L., D.C.P., S.K.C., M.C.S., D.D.R., D.J.H. and D.M.M. designed and performed primary study assays and performed study analysis. J.D.K., A.M.C., J.J.E. and D.M.M. designed, implemented and oversaw the clinical protocol. M.G.H. and R.J.B. designed the data analysis. A.D.K. oversaw pharmacokinetic assays and analysed these data. M.F.K., E.M.A. and J.M.C. designed and performed single-copy assays.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests; details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.M.M. (dmargo@med.unc.edu).

METHODS

Patients/study procedures. HIV-infected patients receiving stable, standard-of-care ART with plasma HIV-1 RNA $<$50 copies per ml and a CD4 count of $>$300 μl^{-1} for at least 6 months were enrolled following informed consent. Studies were approved by the UNC institutional biomedical review board and the Food and Drug Administration. Leukapheresis was performed and *ex vivo* VOR response assayed. Patients with *ex vivo* VOR response were then administered a 200 mg dose of VOR (donated by Merck Research Laboratories) to assess safety and tolerability, followed by a 400 mg dose four or more weeks later during which pharmacokinetic parameters, including maximum concentration (C_{max}) and apparent oral clearance (CL/F), as well as levels of global cellular histone and human p21 promoter histone H3 acetylation. A second 400 mg dose of VOR was administered four or more weeks later, and apheresis spanning the time point of measured C_{max} VOR concentrations was performed to assay *in vivo* VOR response.

Measurement of resting CD4$^+$ T-cell-associated HIV-1 RNA. Leukapheresis and isolation of resting CD4$^+$ T cells have been described elsewhere[4]. To measure the *ex vivo* response of resting CD4$^+$ T cells to VOR, cells were incubated in media containing 335 nM VOR (Merck Research Laboratories) or 3 $\mu g\, ml^{-1}$ PHA (Remel) and 60 U ml^{-1} IL-2 or media alone for 6 h. Cells were washed and plated at 10^6 per well in a 96-well plate and pelleted. Cell pellets were snap frozen and stored at $-80\,^{\circ}C$. To measure the *in vivo* response to VOR, immediately following leukapheresis, resting CD4$^+$ T cells were isolated and plated at 10^6 cells per well, pelleted, snap frozen, and stored at $-80\,^{\circ}C$.

Total RNA was isolated from 12 to 48 pools of 10^6 resting cells using the Magmax 96 Total RNA isolation kit (Ambion) following the manufacturer's protocol. Duplicate pools of cDNA were synthesized from DNase-treated, isolated RNA using the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen) according to the manufacturer's procedures. Two additional duplicate wells from each treatment condition did not include reverse transcriptase and served as control for DNA contamination. Duplicate PCR amplification of duplicate cDNA was performed using the ABI7500 Fast Real-Time PCR machine and previously published primers and probe. A standard curve was generated for each PCR reaction using cDNA synthesized from *in vitro* transcribed RNA where the p5′ plasmid served as template[22]. Results of the four PCR replicates representing each of the original 12 to 48 pools of RNA were averaged and the standard deviation determined for each condition. Inputs for samples in which basal and VOR-induced HIV RNA expression were measured (Fig. 3) were compared by the quantification of TATA binding protein RNA[23] and the quantification of total HIV DNA, and showed no trend towards higher values in the post-VOR samples.

We determined that our HIV RNA PCR assay could detect the difference between 1 copy and $\geq$10 copies using dilutions of an HIV RNA internal standard[24]. Detectable PCR signal less than 10 copies (1–9 copies) was treated in all analyses as 5 copies. No PCR signal or $<$1 copy was treated in all analyses as 0 copies. As the entire pool of cDNA was not amplified, and individual PCR amplifications are therefore subject to stochastic sampling effects, we have termed the result 'relative HIV-1 *gag* RNA copies'. The Wilcoxon rank sum test was used to calculate the statistical significance of all comparisons between conditions.

Measurement of total histone acetylation by flow cytometry. PBMCs collected before and after 400 mg VOR dosing were fixed and permeabilized using Phosflow fix buffer I and Phosflow permeabilization buffer II (BD Biosciences) according to the manufacturer's protocol. Cells were then washed in stain buffer (2% FBS, 0.09% sodium azide), blocked with 8% normal goat serum (Invitrogen), and incubated with anti-acetyl histone H3 (1:100 dilution, catalogue no. 06-599, Millipore) or control rabbit IgG, in blocking solution for 60 min at room temperature. Cells were then washed and incubated with goat-anti-rabbit IgG FITC conjugated secondary antibody (1: 250 dilution, Millipore) in stain buffer for 30 min at 25 °C in the dark. Following a final wash, cells were analysed by flow cytometry using a CyAn ADP flow cytometer and Summit 4.3 software (Beckman Coulter).

Chromatin immunoprecipitation (ChIP). ChIP assays were performed as previously described[9] with the following modifications: PBMCs were crossed-linked with 1% formaldehyde and nuclei extracted using the appropriate buffers (Diagenode). Nuclei were lysed using an SDS lysis buffer containing mammalian protease cocktail inhibitor (Sigma) and sonicated to fragment chromatin to 500–1,000 base pairs using a Bioruptor standard sonicator (Diagenode). 1×10^6 sonicated cells were used to set up each immunoprecipitation reaction using 5 μg of anti-acetyl-histone H3 (Ac-H3, catalogue no. 17-615, Millipore) or rabbit pre-immune immunoglobulin G (Sigma). PCR of immunoprecipitates or input DNA was performed using primers targeting the human p21 promoter[25]. The percent of immunoprecipitated p21 promoter DNA was determined by comparing the cycle threshold values of each reaction to a standard curve generated from input DNA and is reported as the percent of input.

Measurement of VOR pharmacokinetics. Five millilitres of blood were collected without anticoagulant for pharmacokinetic assays before dose and at 0.25, 0.50, 1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0 and 24 h after dosing. Blood samples were allowed to clot at room temperature for 30 min, centrifuged at 2,000*g* for 15 min at 4 °C, and stored at $-70\,^{\circ}C$ until analysis. VOR concentrations were measured from human serum using a modification of the method of ref. 14. Individual serum concentrations were used to estimate VOR pharmacokinetic variables $AUC_{0\text{-}24h}$, C_{max}, T_{max}, CL/F and apparent terminal $t_{1/2}$. WinNonlin Phoenix was used for the calculations. The apparent terminal $t_{1/2}$ was estimated from the best-fit variables of a single exponential to the log-linear portion of the serum concentration versus time curve using unweighted linear regression. $AUC_{0\text{-}24h}$ was calculated using the linear up/log down method. C_{max} and T_{max} were obtained by inspection of the concentration–time data.

24. May, R. C. *et al. Change-Point Models to Estimate the Limit of Detection* (Working Paper 26, University of North Carolina at Chapel Hill Biostatistics Technical Report Series, March 2012); available at http://biostats.bepress.com/uncbiostat/papers/art26.
25. Tonelli, R. *et al.* G1 cell-cycle arrest and apoptosis by histone deacetylase inhibition in MLL-AF9 acute myeloid leukemia cells is p21 dependent and MLL-AF9 independent. *Leukemia* **20,** 1307–1310 (2006).

LETTER

doi:10.1038/nature11288

Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins

Orit Rozenblatt-Rosen[1,2,3]*, Rahul C. Deo[1,4,5]*, Megha Padi[1,6,7]*, Guillaume Adelmant[1,4,8]*, Michael A. Calderwood[1,9,10,11], Thomas Rolland[1,9,10], Miranda Grace[1,12], Amélie Dricot[1,9,10], Manor Askenazi[1,4,8], Maria Tavares[1,2,3,8], Samuel J. Pevzner[9,10,13], Fieda Abderazzaq[1,6], Danielle Byrdsong[1,9,10], Anne-Ruxandra Carvunis[9,10], Alyce A. Chen[1,12], Jingwei Cheng[2,3], Mick Correll[6], Melissa Duarte[1,9,11], Changyu Fan[1,9,10], Mariet C. Feltkamp[14], Scott B. Ficarro[1,4,8], Rachel Franchi[1,9,15], Brijesh K. Garg[1,4,8], Natali Gulbahce[1,9,16,17], Tong Hao[1,9,10], Amy M. Holthaus[1,11], Robert James[1,9,10], Anna Korkhin[1,2,3], Larisa Litovchick[1,2,3], Jessica C. Mar[1,6,7], Theodore R. Pak[18], Sabrina Rabello[1,3,9,16], Renee Rubio[1,6], Yun Shen[1,9,10], Saurav Singh[4,8], Jennifer M. Spangle[1,12], Murat Tasan[1,4,18], Shelly Wanamaker[9,10,15], James T. Webber[4,8], Jennifer Roecklein-Canfield[9,15], Eric Johannsen[1,11], Albert-László Barabási[1,3,9,16], Rameen Beroukhim[3,19,20], Elliott Kieff[1,11], Michael E. Cusick[1,9,10], David E. Hill[1,9,10], Karl Münger[1,12], Jarrod A. Marto[1,4,8], John Quackenbush[1,6,7], Frederick P. Roth[1,4,9,18], James A. DeCaprio[1,2,3] & Marc Vidal[1,9,10]

Genotypic differences greatly influence susceptibility and resistance to disease. Understanding genotype–phenotype relationships requires that phenotypes be viewed as manifestations of network properties, rather than simply as the result of individual genomic variations[1]. Genome sequencing efforts have identified numerous germline mutations, and large numbers of somatic genomic alterations, associated with a predisposition to cancer[2]. However, it remains difficult to distinguish background, or 'passenger', cancer mutations from causal, or 'driver', mutations in these data sets. Human viruses intrinsically depend on their host cell during the course of infection and can elicit pathological phenotypes similar to those arising from mutations[3]. Here we test the hypothesis that genomic variations and tumour viruses may cause cancer through related mechanisms, by systematically examining host interactome and transcriptome network perturbations caused by DNA tumour virus proteins. The resulting integrated viral perturbation data reflects rewiring of the host cell networks, and highlights pathways, such as Notch signalling and apoptosis, that go awry in cancer. We show that systematic analyses of host targets of viral proteins can identify cancer genes with a success rate on a par with their identification through functional genomics and large-scale cataloguing of tumour mutations. Together, these complementary approaches increase the specificity of cancer gene identification. Combining systems-level studies of pathogen-encoded gene products with genomic approaches will facilitate the prioritization of cancer-causing driver genes to advance the understanding of the genetic basis of human cancer.

Integrative studies of viral proteins have identified host perturbations relevant to the aetiology of viral disease[4,5]. We examined whether such a strategy, extended systematically across a range of tumour viruses, could shed light on cancers even beyond those directly caused by these pathogens. Our hypothesis is inspired by classical examples in which DNA tumour virus proteins physically target the products of *RB1* or *TP53*, two well-established germline-inherited and somatically inactivated tumour-suppressor genes[6]. We propose that viruses and genomic variations alter local and global properties of cellular networks in similar ways to cause pathological states. Models derived from host perturbations mediated by viral proteins representing the virome[7] should serve as surrogates for network perturbations that result from large numbers of genomic variations, or the variome[8] (Fig. 1a).

We developed an integrated pipeline to systematically investigate perturbations of host interactome and transcriptome networks induced by gene products of four functionally related, yet biologically distinct, families of DNA tumour viruses: human papillomavirus (HPV), Epstein–Barr virus (EBV), adenovirus (Ad5) and polyomavirus (PyV) (Fig. 1b, Supplementary Table 1 and Supplementary Fig. 1).

Applying a stringent implementation of the yeast two-hybrid (Y2H) system[9], 123 viral open reading frames (viral ORFs) were screened against a collection of about 13,000 human ORFs[10], resulting in a validated virus–host interaction network of 454 binary interactions involving 53 viral proteins and 307 human proteins (Fig. 1c and Supplementary Table 2). Analysis of our binary interaction map identified 31 host target proteins that showed more binary interactions with viral proteins than would be expected given their 'degree' (number of interactors) in our current binary map of the human interactome network (HI-2)[11] (Fig. 1c, Supplementary Table 3 and Supplementary Notes), suggesting a set of common mechanisms by which different viral proteins rewire the host interactome network.

To examine perturbations in both the interactome and transcriptome networks directly in human cells, we generated expression constructs fusing each viral ORF to a tandem epitope tag and introduced each construct into IMR-90 normal human diploid fibroblasts. Interactions between viral proteins and the host proteome were identified by tandem affinity purification followed by mass spectrometry (TAP–MS)[12]. The

[1]Genomic Analysis of Network Perturbations Center of Excellence in Genomic Science, Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [3]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Cardiovascular Research Institute, Department of Medicine and Institute for Human Genetics, University of California, San Francisco, California 94143, USA. [6]Center for Cancer Computational Biology (CCCB), Department of Biostatistics and Computational Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [7]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [8]Blais Proteomics Center and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [9]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [10]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [11]Infectious Diseases Division, The Channing Laboratory, Brigham and Women's Hospital and Departments of Medicine and of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. [12]Division of Infectious Diseases, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [13]Biomedical Engineering Department, Boston University and Boston University School of Medicine, Boston, Massachusetts 02118, USA. [14]Department of Medical Microbiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. [15]Department of Chemistry, Simmons College, Boston, Massachusetts 02115, USA. [16]Center for Complex Networks Research (CCNR) and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA. [17]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, USA. [18]Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto and Lunenfeld Research Institute, Mt Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [19]Department of Medical Oncology and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [20]The Broad Institute, Cambridge, Massachusetts 02142, USA.
*These authors contributed equally to this work.

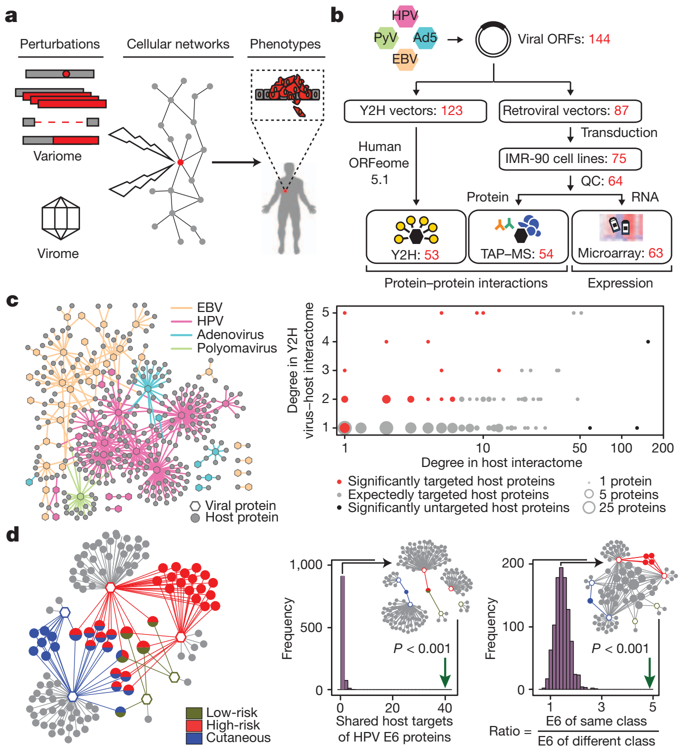


Figure 1 | Systematic mapping of binary interactions and protein complex associations between viral and host proteins. a, The virome-to-variome network model proposes that genomic variations (point mutations, amplifications, deletions or translocations) and expression of tumour virus proteins induce related disease states by similarly influencing properties of cellular networks. **b**, Experimental pipeline for identifying virus–host interactions. Selected cloned viral ORFs were subjected to Y2H screens and introduced into cell lines for both TAP–MS and microarray analyses. Numbers of viral ORFs that were successfully processed at each step are indicated in red. **c**, Left: network of binary virus–host interactions identified by Y2H. Right: subsets of human target proteins that have significantly more (red dots) or fewer (black dots) viral interactors than expected on the basis of their degree in HI-2. **d**, Network of protein complex associations of E6 viral proteins from six HPV types (hexagons, coloured according to disease class) with host proteins (grey circles). Host proteins that associate with two or more E6 proteins are coloured according to the disease class(es) of the corresponding HPV types. Circle size is proportional to the number of associations between host and viral proteins in the E6 networks. Distribution plots of 1,000 randomized networks and experimentally observed data (green arrows) for the number of host proteins targeted by two or more viral proteins in the corresponding sub-networks (left histogram), or the ratio of the probability that a host protein is targeted by viral proteins from the same class to the probability that it is targeted by viral proteins from different classes (right histogram). Insets: representative random networks from these distributions.

intersection of two independent TAP–MS experiments yielded 3,787 reproducibly mapped virus–host protein complex associations involving 54 viral proteins and the products of 1,079 host genes (Supplementary Table 4). Statistically significant overlaps between the Y2H and the TAP–MS data sets with a positive reference set of curated virus–host interactions were observed, supporting the quality of the interactome data sets (Supplementary Notes). Host proteins identified as binary interactors or as members of protein complexes showed a statistically significant overlap ($P < 0.001$) and a statistically significant tendency to interact with each other in HI-2 ($P < 0.001$), implying that host targets in the virus–host interactome maps tend to fall in the same 'neighbourhood' of the host network[1] (Supplementary Fig. 2). Our two complementary interactome data sets highlight specific host biological processes targeted by viral proteins (Supplementary Fig. 3).

To explore the specificity of virus–host relationships, we examined protein complex associations mediated by E6 proteins from six distinct HPV types representing three different disease classes: high-risk mucosal, low-risk mucosal and cutaneous. E6 and E7 proteins encoded by high-risk mucosal HPVs are strongly oncogenic[6]. Multiple host proteins were found to associate with E6 proteins encoded by two or more different HPV types ($P < 0.001$; Fig. 1d), including the known E6 target UBE3A (E6AP)[13]. Among these we observed a statistically significant subgroup of host proteins targeted only by E6 proteins from the same disease class ($P < 0.001$). The transcriptional regulators CREB-binding protein (CREBBP) and EP300 were found to associate with E6 proteins from both cutaneous HPV types, but not with those from the mucosal classes. In contrast with E6 proteins, no group of host proteins showed class-specific targeting by HPV E7 proteins (Supplementary Fig. 4). These differential associations reflect how rewiring of virus–host interactome networks may relate to the aetiology of viral disease.

In addition to targeting protein–protein interactions, viral proteins also functionally perturb their hosts through downstream effects on gene expression. We profiled the transcriptome of the viral ORF-transduced cell lines to trace pathways through which viral proteins could alter cellular states. Model-based clustering of the 2,944 most frequently perturbed host genes identified 31 clusters, many of which were enriched ($P < 0.01$) for specific Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Fig. 2a and Supplementary Tables 5 and 6). To uncover transcription factor (TF) binding motifs enriched within the promoters or enhancers of the corresponding genes, a high-confidence map of predicted TF binding sites was generated by using information on cell-specific chromatin accessibility and consensus TF-binding motifs (Supplementary Fig. 5). We found a densely interconnected set of 92 TFs (Supplementary Fig. 6 and Supplementary Notes) that either associated with or were differentially expressed in response to viral proteins, and for which target genes were enriched in at least one cluster (Supplementary Fig. 7 and Supplementary Table 7).

The mean expression change of each cluster revealed three distinct groups of viral proteins (Fig. 2a): group I included low-risk and cutaneous HPV E6 proteins, group II contained most of the EBV proteins, and group III included high-risk HPV E6 and E7 proteins and polyomavirus proteins. Consistent with their ability to associate with RB1, group III viral proteins increased the expression of genes that are involved in cell proliferation and whose promoters are enriched in E2F binding sites (clusters C26 and C31). Steady-state levels of these genes are correlated with cellular growth phenotypes (Supplementary Fig. 8 and Supplementary Notes). Similarly, the decreased expression of the p53 signalling pathway probably reflects the ability of group III proteins to bind to and inactivate p53 (cluster C12).

To investigate additional pathways through which viral proteins perturb TFs to reprogram cellular states, we derived a detailed network model containing 58 viral proteins that perturb the activity of 86 TFs, which in turn potentially regulate 30 clusters (Supplementary Fig. 9). This model was predictive of downstream patterns of differential expression ($P = 0.003$; Fig. 2b) and suggested ways in which viral proteins could regulate many of the biological hallmarks of cancer[14] (Supplementary Notes). For example, we found the regulation of several pathways involved in the response to DNA damage (Fig. 2b), including autophagy potentially through NFE2L2 (cluster C3)[15], the NF-κB-mediated inflammatory response (cluster C23)[16], and the type I interferon response through IRF-1 (cluster C24; Supplementary Fig. 5)[17].

Specific disease outcomes of the three disease classes of HPV might reflect how their respective proteins perturb distinct functional groups of host proteins. Our protein complex interaction map revealed associations between E6 from cutaneous HPVs and Mastermind-like protein 1 (MAML1), EP300 and CREBBP. MAML1 forms a transcriptional activation complex that modulates the expression of Notch target genes in conjunction with the EP300 and CREBBP histone acetyltransferases, the RBPJ transcription factor and the intracellular domain (ICD) of the Notch receptor[18]. Our transcriptome profiling placed the cutaneous

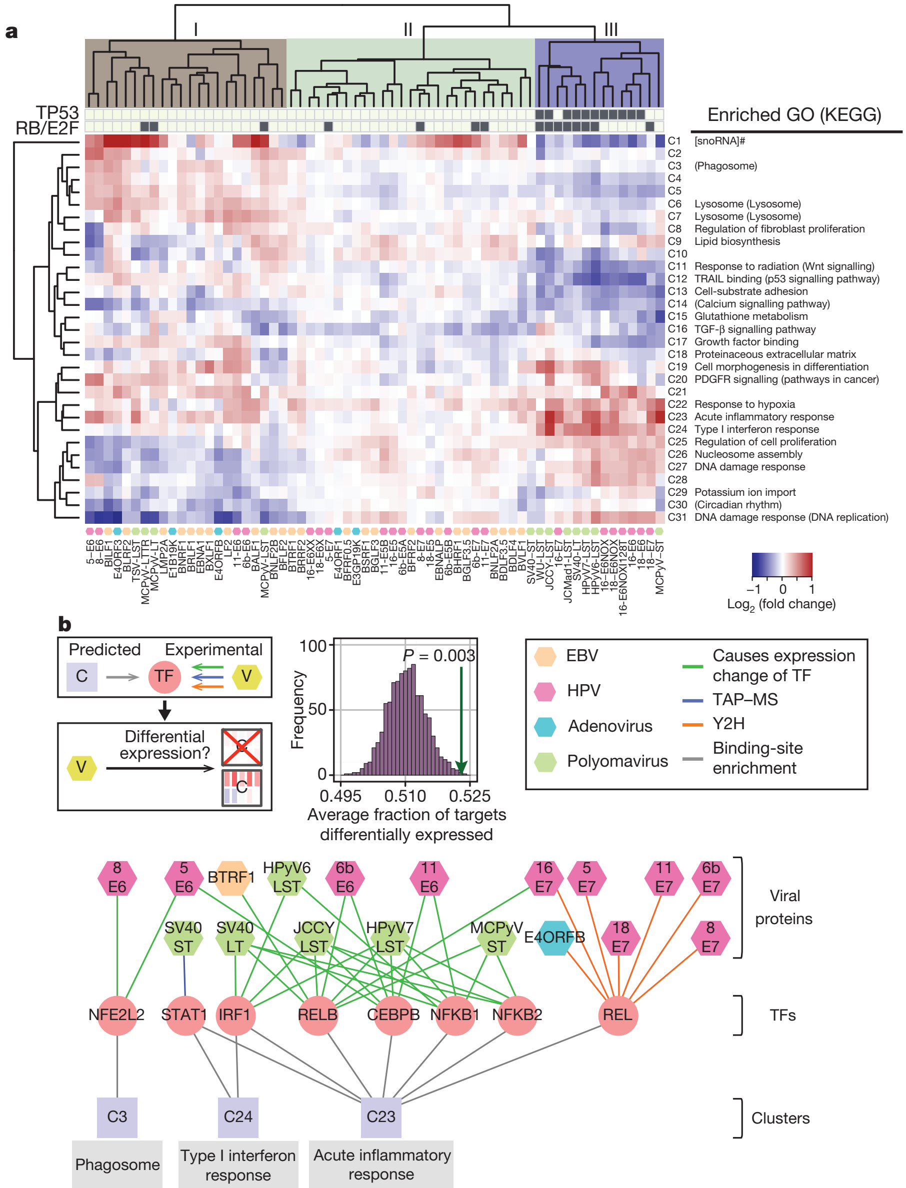


Figure 2 | Transcriptome perturbations induced by viral protein expression. **a**, Heat map of average cluster expression relative to control. Enriched GO terms and KEGG pathways are listed adjacent to the numbered expression clusters. In cluster C1, eight of the nine transcripts are small nucleolar RNAs (denoted by hash mark). The upper dendrogram is shaded by viral ORF grouping. Grey blocks show which viral proteins associate with the indicated host proteins. **b**, Diagram showing how the viral protein–TF–target-gene network was constructed, with paths to three clusters shown. C, cluster; V, viral protein. Null distribution of average fraction of TF target genes differentially expressed in the corresponding cell lines (histogram), along with the observed value (green arrow).

and low-risk HPV E6 proteins in group I apart from the high-risk HPV E6 proteins in group III (Fig. 2a), so we investigated these differential perturbations. Both cutaneous HPV5 and HPV8 E6 proteins co-precipitated MAML1 and EP300, whereas the mucosal HPV E6 proteins did not (Fig. 3a and Supplementary Fig. 10); conversely, HPV6b, HPV11, HPV16 and HPV18 E6 proteins associated with UBE3A[13], whereas cutaneous HPV E6 proteins did not.

Perturbations in Notch signalling can confer either oncogenic or tumour-suppressive effects[18]. Because both inhibition of the Notch pathway and the expression of HPV8 E6 promote squamous cell carcinoma[19,20], we reasoned that binding of HPV5 and HPV8 E6 to MAML1 might inhibit Notch signalling. To test this, we examined transcript levels of Notch pathway genes and potential Notch target genes with a predicted RBPJ binding site in their promoter across all HPV E6 cell lines as well as in cells depleted for MAML1. Transcript levels of several Notch targets were significantly decreased in IMR-90 cells that were either depleted for MAML1 or expressing either HPV5 or HPV8 E6 (Fig. 3b and Supplementary Fig. 11). These and other results[21] indicate that the association of HPV5 and HPV8 E6 proteins with MAML1 inhibits Notch signalling. Building on these observations and on the established associations between EBV EBNA proteins and RBPJ[22], we observed that viral proteins from all four DNA tumour viruses target proteins of the Notch pathway ($P < 0.002$; Fig. 3c). Our data highlight the central role of Notch signalling in both virus–host perturbations and tumorigenesis, and support observations that implicate *MAML1* in cancer pathogenesis[23].

We next investigated the extent to which viral proteins globally target host proteins causally implicated in cancer. First we compared our viral targets, identified through binary interaction, protein complex associations and TF-binding-site analyses, against a gold standard set of 107 high-confidence causal human cancer genes in the COSMIC Classic (CC) gene set[24] (Supplementary Table 8). Viral targets were

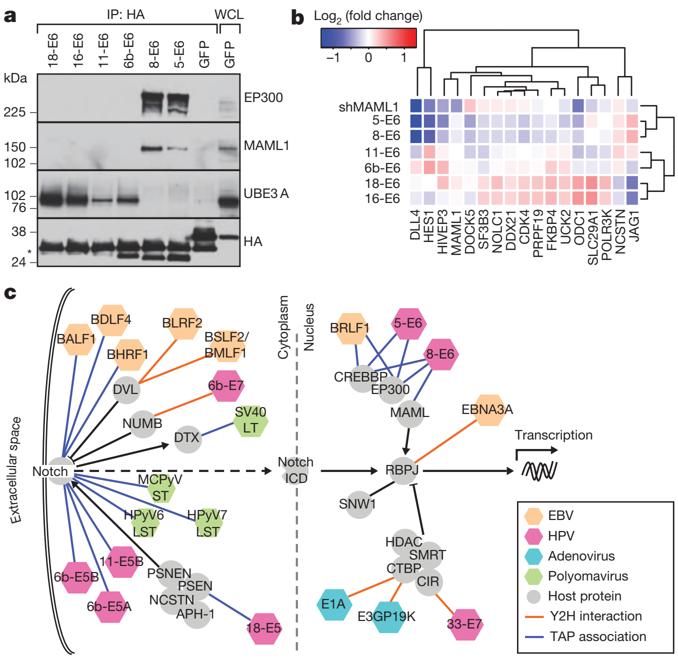


Figure 3 | The Notch pathway is targeted by multiple DNA tumour virus proteins. **a**, Western blots of co-immunoprecipitations of HPV E6 proteins in IMR-90 cells. **b**, Heat map of expression of Notch-pathway-responsive genes in IMR-90 cells on expression of E6 proteins from different HPV types or on knockdown of *MAML1*, relative to control cells. **c**, Representation of viral protein interactions with components of the Notch signalling pathway (as defined in KEGG). Notch ICD, Notch intracellular domain.

significantly enriched among CC genes (Supplementary Fig. 12; adjusted P for multiple hypothesis testing (P_{adj}) = 0.01). To optimize the stringency of potential cancer enrichment analyses, we restricted the set of viral protein targets identified by TAP–MS to those identified by three or more unique peptides, a choice corresponding to an experimental reproducibility rate greater than 90% (Supplementary Fig. 13). The resulting stringent candidate set of 947 host target genes (the 'VirHost' set; Supplementary Table 9) included 16 proteins encoded by CC genes ($P = 0.007$; Fig. 4a), among which tumour suppressor genes were significantly over-represented ($P = 0.03$).

As a complementary approach to validate our VirHost gene set, we compiled a list of human orthologues of mouse genes recently implicated in tumorigenesis by *in vivo* transposon mutagenesis screens[25]. Our VirHost data set overlaps significantly with these candidate cancer genes ($P < 0.0001$) (Fig. 4b and Supplementary Table 10). The 156 candidate genes in the overlap were markedly enriched both for CC genes (odds ratio (OR) = 13, $P = 4 \times 10^{-9}$) and for genes implicated in apoptosis, hypoxia response and cell growth pathways ($P_{adj} < 0.05$ for all). Taken together, these observations suggest that our VirHost data set points to previously unknown human cancer-associated genes.

Large-scale tumour-sequencing efforts have the potential to discover new tumour suppressors and oncogenes. To explore how the VirHost data set might be used to interpret these data, we compiled somatic mutations for eight different cancers identified through 12 sequencing projects. Non-synonymous somatic mutations were reported for 10,543 genes, so further prioritization is required for useful identification of candidate causal cancer genes. We therefore scored the likely functional effects of these mutations by using the PolyPhen2 program[26] and generated a cumulative somatic mutation (SM) score for each protein (Supplementary Fig. 14). To compare performance in identifying candidate cancer genes of our VirHost set with that of proteins ranked by SM analysis, we tested a matching number (947) of the top-ranked SM candidates for overlap with CC genes (Fig. 4c and Supplementary Table 11). In comparison with the 16 cancer genes identified in our VirHost set, SM recovered 23 genes

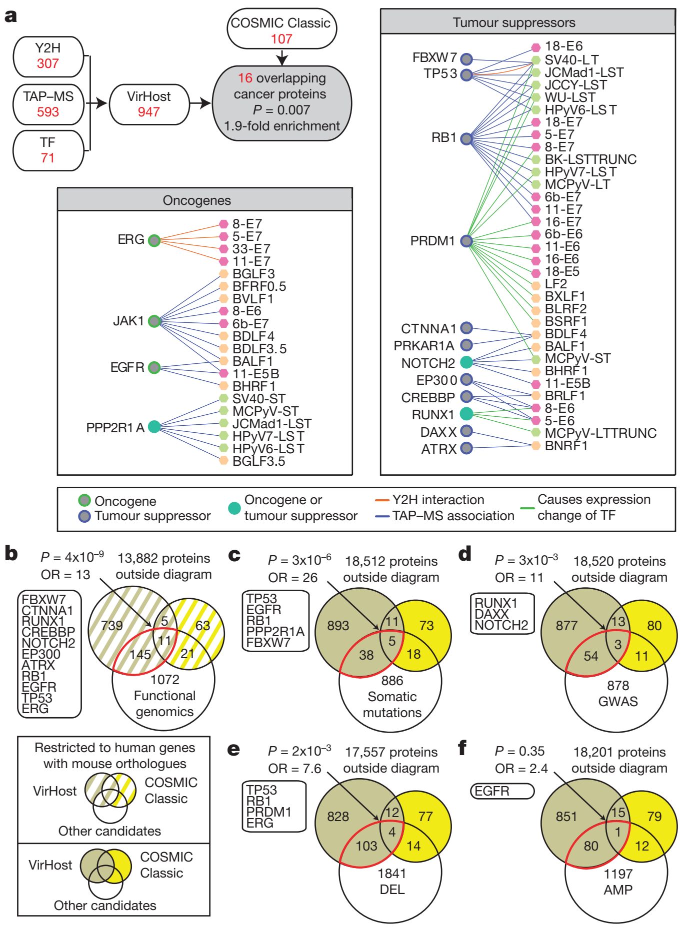


Figure 4 | Interpreting cancer genomes with the use of virus–host network models. **a**, Diagram describing the composition of VirHost (proteins identified by TAP–MS with at least three unique peptides, Y2H and TF) and overlap with COSMIC Classic (CC) genes. Viral protein (hexagons) perturbations of cancer proteins (circles), with cancer proteins classified as oncogenes or tumour suppressors. **b**, Venn diagram of overlaps of VirHost proteins with CC genes and candidate cancer genes identified through four transposon-based functional genomics screens. **c**, Venn diagram of overlaps of VirHost proteins with CC genes and with a prioritized set of genes found through somatic mutation analysis. P values are based on Fisher's exact test or permutation. **d**–**f**, Venn diagrams comparing VirHost, GWAS (**d**), SCNA-DEL (**e**) and SCNA-AMP (**f**) data sets for ability to recover COSMIC Classic genes.

($P = 6 \times 10^{-10}$). Thus, viral perturbation analysis is comparable to somatic mutation sequencing for identifying cancer genes.

Although both strategies showed significant overlap with the reference CC gene set, neither by itself suffices to pinpoint causal genes with high specificity. To overcome this difficulty we exploited the orthogonal nature of the VirHost and SM sets (given $P = 0.58$ for their overlap) by focusing on the 43 proteins at their intersection (the 'VirHostSM' subset) (Fig. 4c). In contrast with VirHost (OR = 3.7) or SM (OR = 5.8), the VirHostSM set was markedly enriched in CC proteins (five proteins, OR = 26, $P = 3 \times 10^{-6}$). Pathway analysis of the 43 proteins revealed 12 proteins implicated in the GO pathway linked to 'regulation of apoptosis' (OR = 6.0, $P_{adj} = 0.017$). The intersection also includes plausible contributors to cancer pathogenesis (Supplementary Fig. 15) such as the oxidative stress response transcription factor NFE2L2.

We compared the ability of VirHost to identify CC genes with two other large-scale genomic approaches: SCNA (somatic copy number alteration)[27] analysis of cancer genomes and GWAS (genome-wide association studies) of cancer susceptibility[28]. The SCNA deletions (SCNA-DEL) and amplifications (SCNA-AMP) and GWAS sets all significantly overlapped with CC genes, but with lower specificity than

the VirHost overlap with CC (OR = 1.9 for SCNA-DEL, 2.1 for SCNA-AMP, and 3.1 for GWAS, versus 3.7 for VirHost; Fig. 4d–f). The intersections of VirHost with GWAS or SCNA-DEL genes also showed enrichment for cancer genes (Supplementary Table 12). The intersection of VirHost and SCNA-DEL was enriched for genes implicated in apoptosis (GO term 'programmed cell death'; 15 genes, $P_{adj} = 0.022$, OR = 4.3). Conversely, there was no synergy in the intersection of SCNA-AMP and VirHost, perhaps reflecting the preference of viral proteins in targeting tumour suppressors rather than oncogenes (Fig. 4f).

Our systems-level explorations of viral perturbations facilitate the distinction between driver and passenger mutations in cancer genome sequences. Our data indicate that *trans*-acting viral products and *cis*-acting genome variations involved in cancer converge on common pathways.

METHODS SUMMARY

Viral ORF entry clones were generated by PCR-based Gateway recombinational cloning[4]. After sequence verification, viral ORFs were transferred by *in vitro* Gateway LR recombinational cloning into expression vectors for Y2H screening[9] and for transduction of IMR-90 cells. Y2H screens were performed against the human ORFeome v5.1 collection of about 13,000 full-length human ORFs[10]. Total RNA was isolated from IMR-90 cells expressing viral ORFs, and gene expression was assayed on Human Gene 1.0 ST arrays. Microarray data were analysed with R/Bioconductor. Viral proteins and associated host proteins were purified by sequential Flag and haemagglutinin immunoprecipitation and analysed by liquid chromatography–tandem mass spectrometry. Virus–host protein complex associations from two independent purifications were analysed. Pathway enrichment was analysed with FuncAssociate[29]. Assessment of statistical significance for overlap between gene sets was performed with Fisher's exact test or resampling-based approaches.

Received 8 June 2011; accepted 7 June 2012.
Published online 18 July 2012.

1. Vidal, M., Cusick, M. E. & Barabási, A. L. Interactome networks and human disease. *Cell* **144,** 986–998 (2011).
2. Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331,** 1553–1558 (2011).
3. Gulbahce, N. *et al.* Viral perturbations of host networks reflect disease etiology. *PLOS Comput. Biol.* **8,** e1002531 (2012).
4. Calderwood, M. A. *et al.* Epstein–Barr virus and virus human protein interaction maps. *Proc. Natl Acad. Sci. USA* **104,** 7606–7611 (2007).
5. Shapira, S. D. *et al.* A physical and regulatory map of host–influenza interactions reveals pathways in H1N1 infection. *Cell* **139,** 1255–1267 (2009).
6. Howley, P. M. & Livingston, D. M. Small DNA tumor viruses: large contributors to biomedical sciences. *Virology* **384,** 256–259 (2009).
7. Foxman, E. F. & Iwasaki, A. Genome–virome interactions: examining the role of common viral infections in complex disease. *Nature Rev. Microbiol.* **9,** 254–264 (2011).
8. Editorial. What is the Human Variome Project? *Nature Genet.* **39,** 423 (2007).
9. Dreze, M. *et al.* High-quality binary interactome mapping. *Methods Enzymol.* **470,** 281–315 (2010).
10. Lamesch, P. *et al.* hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* **89,** 307–315 (2007).
11. Yu, H. *et al.* Next-generation sequencing to generate interactome datasets. *Nature Methods* **8,** 478–480 (2011).
12. Zhou, F. *et al.* Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. *J. Proteome Res.* **9,** 6242–6255 (2010).
13. Brimer, N., Lyons, C. & Vande Pol, S. B. Association of E6AP (UBE3A) with human papillomavirus type 11 E6 protein. *Virology* **358,** 303–310 (2007).
14. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).
15. Fujita, K., Maeda, D., Xiao, Q. & Srinivasula, S. M. Nrf2-mediated induction of p62 controls Toll-like receptor-4-driven aggresome-like induced structure formation and autophagic degradation. *Proc. Natl Acad. Sci. USA* **108,** 1427–1432 (2011).
16. Wu, Z. H., Shi, Y., Tibbetts, R. S. & Miyamoto, S. Molecular linkage between the kinase ATM and NFκB signaling in response to genotoxic stimuli. *Science* **311,** 1141–1146 (2006).
17. Tanaka, N. *et al.* Cooperation of the tumour suppressors IRF-1 and p53 in response to DNA damage. *Nature* **382,** 816–818 (1996).
18. Ranganathan, P., Weaver, K. L. & Capobianco, A. J. Notch signalling in solid tumours: a little bit of everything but not all the time. *Nature Rev. Cancer* **11,** 338–351 (2011).
19. Proweller, A. *et al.* Impaired Notch signaling promotes *de novo* squamous cell carcinoma formation. *Cancer Res.* **66,** 7438–7444 (2006).
20. Marcuzzi, G. P. *et al.* Spontaneous tumour development in human papillomavirus type 8 E6 transgenic mice and rapid induction by UV-light exposure and wounding. *J. Gen. Virol.* **90,** 2855–2864 (2009).
21. Brimer, N., Lyons, C., Wallberg, A. E. & Vande Pol, S. B. Cutaneous papillomavirus E6 oncoproteins associate with MAML1 to repress transactivation and NOTCH signaling. *Oncogene*. doi:10.1038/onc.2011.589 (16 January 2012).
22. Calderwood, M. A. *et al.* Epstein–Barr virus nuclear protein 3C binds to the N-terminal (NTD) and beta trefoil domains (BTD) of RBP/CSL; only the NTD interaction is essential for lymphoblastoid cell growth. *Virology* **414,** 19–25 (2011).
23. Klinakis, A. *et al.* A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. *Nature* **473,** 230–233 (2011).
24. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950 (2011).
25. Copeland, N. G. & Jenkins, N. A. Harnessing transposons for cancer gene discovery. *Nature Rev. Cancer* **10,** 696–706 (2010).
26. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7,** 248–249 (2010).
27. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463,** 899–905 (2010).
28. Manolio, T. A. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363,** 166–176 (2010).
29. Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19,** 2502–2504 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Center for Cancer Systems Biology (CCSB) and J. Aster, M. Meyerson, W. Kaelin, G. Superti-Furga and S. Sunyaev for discussions, and J.W. Harper, W. Hahn, P. Howley, Y. Jacob, M. Imperiale, I. Koralnik, H. Pfister and D. Wang for reagents. This work was primarily supported by Center of Excellence in Genomic Science (CEGS) grant P50HG004233 from the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) awarded to M.V. (principal investigator), A.-L.B., J.A.D., E.K., J.A.M., K.M., J.Q. and F.P.R. Additional funding included Institute Sponsored Research funds from the Dana-Farber Cancer Institute Strategic Initiative to M.V.; NIH grants R01HG001715 to M.V., D.E.H. and F.P.R.; R01CA093804, R01CA063113 and P01CA050661 to J.A.D.; R01CA081135, R01CA066980 and U01CA141583 to K.M.; R01CA131354, R01CA047006 and R01CA085180 to E.K.; T32HL007208 and K08HL098361 to R.C.D.; K08CA122833 to R.B.; F32GM095284 and K25HG006031 to M.P.; Canada Excellence Research Chairs (CERC) Program, Canadian Institute for Advanced Research Fellowship and Ontario Research Fund to F.P.R.; and James S. McDonnell Foundation grant 220020084 to A.-L.B. M.V. is a 'Chercheur Qualifié Honoraire' from the Fonds de la Recherche Scientifique (Wallonia-Brussels Federation, Belgium).

Author Contributions O.R.-R., G.A., M.A.C., M.G., A.D., Ma.T., F.A., D.B., A.A.C., J.C., M.C., M.D., M.C.F., S.B.F., R.F., B.K.G., A.M.H., R.J., A.K., L.L., R.R., J.M.S., S.W., J.R.-C. and E.J. performed experiments or contributed new reagents. R.C.D., M.P., G.A., T.R., M.A., S.P., A.-R.C., C.F., N.G., T.H., J.C.M., T.R.P., S.R., Y.S., S.S., Mu.T. and J.T.W. performed computational analysis. O.R.-R., R.C.D., M.P., G.A., M.A.C., T.R., M.E.C., D.E.H., K.M., J.A.M., F.P.R., J.A.D. and M.V. wrote the manuscript. A.-L.B., R.B., E.K., M.E.C., D.E.H., K.M., J.A.M., J.Q., F.P.R., J.A.D. and M.V. designed or advised research. M.A.C., T.R., M.G., A.D., M.A., Ma.T. and S.P. contributed equally and should be considered joint second authors; D.E.H., K.M., J.A.M., J.Q., F.P.R., J.A.D. and M.V should be considered joint senior authors; additional co-authors are listed alphabetically.

Author Information Microarray data were deposited in the Gene Expression Omnibus database under accession number GSE38467. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.M. (jarrod_marto@dfci.harvard.edu), J.Q. (johnq@jimmy.harvard.edu), F.P.R. (fritz.roth@utoronto.ca), J.A.D. (james_decaprio@dfci.harvard.edu) and M.V. (marc_vidal@dfci.harvard.edu).

doi:10.1038/nature11289

Viral immune modulators perturb the human molecular network by common and unique strategies

Andreas Pichlmair[1,2], Kumaran Kandasamy[1], Gualtiero Alvisi[3], Orla Mulhern[4], Roberto Sacco[1], Matthias Habjan[2,5], Marco Binder[3], Adrijana Stefanovic[1], Carol-Ann Eberle[1], Adriana Goncalves[1], Tilmann Bürckstümmer[1], André C. Müller[1], Astrid Fauster[1], Cathleen Holze[2], Kristina Lindsten[6], Stephen Goodbourn[7], Georg Kochs[5], Friedemann Weber[5,8,9], Ralf Bartenschlager[3], Andrew G. Bowie[4], Keiryn L. Bennett[1], Jacques Colinge[1] & Giulio Superti-Furga[1]

Viruses must enter host cells to replicate, assemble and propagate. Because of the restricted size of their genomes, viruses have had to evolve efficient ways of exploiting host cell processes to promote their own life cycles and also to escape host immune defence mechanisms[1,2]. Many viral open reading frames (viORFs) with immune-modulating functions essential for productive viral growth have been identified across a range of viral classes[3,4]. However, there has been no comprehensive study to identify the host factors with which these viORFs interact for a global perspective of viral perturbation strategies[5–11]. Here we show that different viral perturbation patterns of the host molecular defence network can be deduced from a mass-spectrometry-based host-factor survey in a defined human cellular system by using 70 innate immune-modulating viORFs from 30 viral species. The 579 host proteins targeted by the viORFs mapped to an unexpectedly large number of signalling pathways and cellular processes, suggesting yet unknown mechanisms of antiviral immunity. We further experimentally verified the targets heterogeneous nuclear ribonucleoprotein U, phosphatidylinositol-3-OH kinase, the WNK (with-no-lysine) kinase family and USP19 (ubiquitin-specific peptidase 19) as vulnerable nodes in the host cellular defence system. Evaluation of the impact of viral immune modulators on the host molecular network revealed perturbation strategies used by individual viruses and by viral classes. Our data are also valuable for the design of broad and specific antiviral therapies.

We performed a survey to identify the cellular proteins and associated complexes interacting with 70 viORFs inducibly expressed from an identical genomic locus in a human cell line (HEK293 Flp-In TREx) competent for innate antiviral programs[12,13](Fig. 1a). This set-up allowed us to gauge the expression levels of the viral proteins and to assess the formation of endogenous protein complexes under physiological conditions in human cells[14]. We selected the viORFs to cover four groups of viruses representative of ten different families and checked for their correct expression (Supplementary Figs 1, 2a–c and 3 and Supplementary Table 1)[15] and, in selected cases, immune modulatory activity (Supplementary Fig. 2d, e)[16,17]. We isolated interacting cellular proteins by tandem affinity purification (TAP) and analysed purified proteins by one-dimensional gel-free liquid chromatography tandem mass spectrometry (LC–MS/MS) (Supplementary Fig. 4a, b)[18]. The 70 viORFs specifically interacted with 579 cellular proteins with high confidence, resulting in 1,681 interactions (Fig. 1a, Supplementary Fig. 4c and Supplementary Table 1; see Methods for details). To validate our approach we assessed the impact of viral infection on the identified viORF–host-protein interactions with the use of several cognate viruses and found decreased numbers of co-purifying proteins, probably as a result of decreased cellular viability as well as competition with the tagged viORF (Supplementary Fig. 5). In addition, treatment with type I interferon (IFN) (Supplementary Fig. 4d) to simulate a host immune response had little effect on the interaction pattern of selected viORFs (Supplementary Fig. 5).

Of the 579 cellular proteins identified as interacting with the 70 viORFs, there was a strong enrichment for proteins associated with innate immunity, further validating the approach and potentially revealing additional unknown components of the host antiviral defence network (overlap with InnateDB database[19]; $P < 2.3 \times 10^{-47}$) (Supplementary Fig. 6a and Supplementary Table 2). There was also a strong enrichment for ubiquitously expressed proteins[20] ($P < 2.2 \times 10^{-138}$) and for evolutionarily conserved proteins ($P < 2.2 \times 10^{-16}$) consistent with the coevolution of virus–host relationships (Supplementary Fig. 6b–d and Supplementary Table 3).

To obtain a more comprehensive view of how viORFs influence host cell processes, we used quantitative information from the mass spectrometry data to compute the strength of impact of each viORF on its cellular targets, and used these quantitative parameters in all subsequent analyses. We also incorporated data from the human protein–protein interactome (humPPI) assembled from public databases, to analyse the protein network associated with the viORF-interacting cellular targets. We found that in comparison with an average human

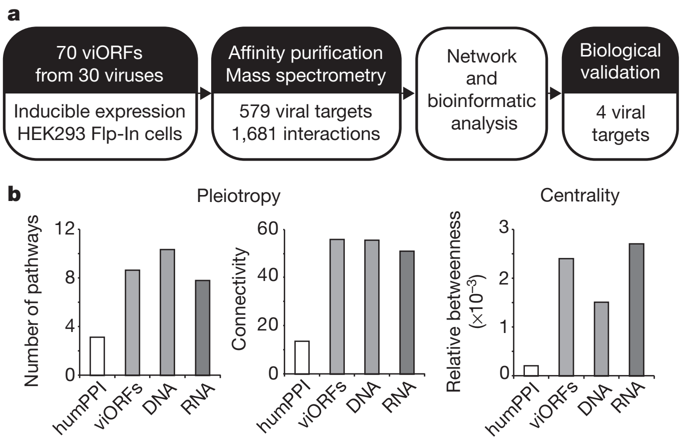


Figure 1 | Host factor survey set-up and general properties of the data set. **a**, Workflow of the host factor survey. **b**, Topological network properties of proteins identified as targets of viral proteins. The histograms compare the average property of proteins in the humPPI with the entire group of viORF interactors, or with viORFs derived from viruses with DNA and RNA genomes, respectively.

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria. [2]Innate Immunity Laboratory, Max Planck Institute of Biochemistry, 82152 Martinsried/Munich, Germany. [3]Department of Infectious Diseases, Molecular Virology, Heidelberg University, 69120 Heidelberg, Germany. [4]School of Biochemistry and Immunology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin 2, Ireland. [5]Department of Virology, University of Freiburg, 79104 Freiburg, Germany. [6]Department of Cell and Molecular Biology, Karolinska Institutet, 17177 Stockholm, Sweden. [7]Division of Basic Medical Sciences, St George's, University of London, London SW17 0RE, UK. [8]Centre for Biological Signalling Studies (BIOSS), Albert-Ludwigs-Universität Freiburg, 79108 Freiburg, Germany. [9]Institute for Virology, Philipps-University Marburg, 35043 Marburg, Germany.

protein, the average viral target was distinct in four ways: it was significantly more connected to other proteins; it was in a more central network position; it participated in more cellular pathways; and it was more likely to be engaged in central positions within these pathways (Fig. 1b and Supplementary Fig. 6d, e). These properties are consistent with a strong influence on pathways and effective control of biological networks[21], which is in line with the parsimonious use of viral genetic material, and coevolution of the virus with the host organism.

Our large host-factor survey using a defined cellular set-up offers the unique opportunity to identify host-cell perturbation strategies pursued by individual viruses, families and groups. On the basis of the humPPI, 70% of the viORF-interacting cellular factors formed a coherent protein–protein interaction network (Supplementary Fig. 7a). When mapped on the entire humPPI, viral targets seemed to occupy central positions (Supplementary Fig. 7b). We also grouped the cellular targets on the basis of their interaction with viORFs from single-stranded (ss) or double-stranded (ds) RNA or DNA viruses and found that about half of the viORF targets linked to a single viral group, and the rest interacted with viruses of more than one group (Fig. 2a). Statistically significant enrichment for individual gene ontology (GO) terms, representing categories of biological processes, could be identified for each subnetwork. Proteins targeted by ssRNA(−) viORFs were enriched for processes related to protection of the viral genome and transcripts from degradation or detection by the host, and for those promoting efficient viral RNA processing (Fig. 2a). This is illustrated by the interaction between NS1 of influenza A virus (FluAV) with the 5′→3′ exoribonuclease XRN2, and among the NSs protein of Rift Valley fever virus, the mRNA export factor RAE1 and the nuclear pore complex protein NUP98. In contrast, dsRNA virus targets were enriched for protein catabolic processes (Fig. 2a) with both rotaviruses and reoviruses (NSP1 and σ3) engaging SKP1–CUL1–F-box protein complexes (containing FBXW11, Cullin-3, and Cullin-7 and Cullin-9, respectively), which mediate protein degradation.

To determine which cellular signalling pathways are targeted by viORFs and to look for differences between DNA and RNA viruses, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations (Supplementary Table 4). Clear distinctions in preferences were observed between the different viral groups, with viORFs of RNA viruses targeting the JAK–STAT and chemokine signalling pathways, as well as pathways associated with intracellular parasitism, and viORFs of DNA viruses targeting cancer pathways (glioma, acute myeloid leukaemia and prostate cancer) (Supplementary Table 4). Among the viral targets that are involved in multiple cellular pathways were two catalytic and three regulatory subunits of the phosphatidylinositol-3-OH kinase family, identified with the FluAV NS1 protein and with the TLR inhibitory protein A52 of vaccinia virus (VACV) (Supplementary Fig. 8a)[4]. We functionally validated these interactions and identified a critical role for one of the catalytic subunits (PIK3CA) in TRIF-mediated IFN-β promoter activation (Supplementary Fig. 8b–d).

The higher probability of viORFs targeting cellular proteins that link different pathways (Fig. 1b and Supplementary Fig. 6d) prompted us to map which of these pathway connections were preferentially targeted and thus were probably disrupted (Fig. 2b), and to compare the disruption patterns brought about by viORFs from DNA viruses with

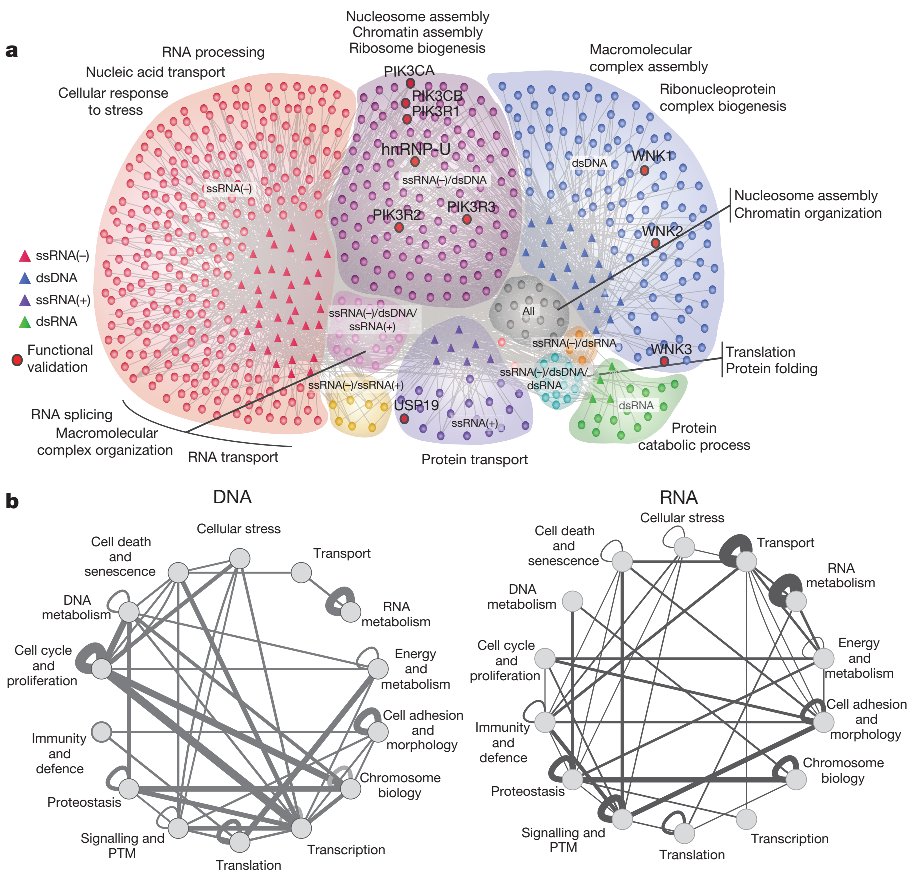


Figure 2 | Network of identified targets and network perturbation induced by viORFs. **a**, Network representation of all the viORF–target-protein interactions with viral targets grouped according to the genome type of the interacting viORF(s). Proteins identified in the negative control cell line were subtracted as non-specific binders. Triangles represent viORFs; circles represent viral target proteins. Protein interactions functionally validated in detail in the study are marked in dark red. Up to three GO terms significantly enriched in the corresponding viral target subsets are shown around the network to highlight specific functions. **b**, viORFs targeting one or two proteins that physically interact and are involved in one or more biological processes have the potential to perturb communication or synchronization within or between the given process(es). Significant perturbations were determined ($P < 0.001$) using targets of viORFs derived from DNA or RNA viruses; edge thickness represents a normalized perturbation score.

those from RNA viruses. About one-third of the connections between specific cellular processes were hit by both viral types, suggesting a similar mechanism of perturbing the host cells. viORFs from DNA viruses preferentially targeted proteins linking the cell cycle with either transcription or chromosome biology, possibly reflecting the necessity of uncoupling viral replication from cellular growth. In contrast, RNA viruses targeted proteins involved in RNA metabolism and also protein and RNA transport, while preferentially disrupting the link between signalling and immunity-related processes (Fig. 2b).

To integrate our viORF–host-protein interaction data sets with intracellular events occurring after viral infection we compared our viORF interaction proteomic profile with the transcriptional profile obtained after infection of the cells with hepatitis C virus (HCV) (Supplementary Table 5). The protein-processing pathway in the endoplasmic reticulum (ER) (Supplementary Fig. 9a) was the most affected process. The HCV viORFs specifically targeted six ER-associated proteins. To analyse the broader implications of this targeting on the cell, we identified the cellular proteins known to bind to these six ER targets and analysed their functions bioinformatically (Supplementary Fig. 9b). Of the 80 cellular protein interactors, 42 were enriched in either cell-cycle or apoptosis functions (Supplementary Fig. 9c). Ubiquitin-specific peptidase 19 (USP19), a deubiquitinating enzyme involved in the unfolded protein response[22], interacted with the viORF NS5A. To study the biological relevance of this interaction, we analysed the localization of USP19 after HCV infection and found that it relocalized to HCV replication compartments in replicon-containing cells, probably disrupting its cellular function (Supplementary Fig. 10a, b). Indeed, NS5A inhibited the ability of USP19 to rescue destabilized green fluorescent protein (GFP) that was degraded by the proteasome (Fig. 3a). In addition, infection of cells with wild-type HCV decreased cell growth[23], whereas infection with recombinant virus lacking the NS5A–USP19 interaction site, which mapped to 50 amino acids in domain III (Supplementary Fig. 10c–g), did not (Fig. 3b and Supplementary Fig. 10h). Thus, the cell-proliferation-inhibitory properties of NS5A are probably mediated by its inhibition of USP19, which is known to promote cell growth[24], and implicates the targeting of ER-resident proteins and proteostasis as an important viral perturbation strategy.

The heterogeneous ribonucleoprotein hnRNP-U was among the most frequently targeted cellular proteins in the analysis (Supplementary Figs 11 and 12a and Supplementary Table 6) and has previously been reported to restrict growth of HIV[25]. Overexpression of hnRNP-U inhibited the polymerase activity of FluAV and the growth of vesicular stomatitis virus (VSV) (Supplementary Fig. 12b and data not shown). This inhibitory effect was alleviated by coexpression of NS1 (FluAV), establishing a functional link to hnRNP-U (Fig. 3c). We mapped the NS1 interaction site on hnRNP-U to the carboxy-terminal Arg-Gly-Gly (RGG) domain (Fig. 3d and Supplementary Fig. 12c)[26]. The RGG domain bound viral RNA in infected cells (Supplementary Fig. 12d), and an hnRNP-U mutant lacking this domain was defective in antiviral polymerase inhibition (Fig. 3e), suggesting that hnRNP-U inhibits the replication of RNA-viruses through viral RNA interaction. Collectively, the analysis highlights hnRNP-U as an important antiviral protein and a hotspot of viral perturbation strategies.

Of the 70 viORFs used in the study, only K7 of VACV[27] interacted with members of the WNK family (Supplementary Figs 11 and 13a–e and Supplementary Table 6), which are regulators of ion transport and are implicated in cancer[28]. Subsequent analyses on the potential role of this protein family in the antiviral immune response revealed that WNK1 and WNK3, but not WNK2 or WNK4, synergized with interleukin-1 (IL-1)-stimulated activation of the p38 kinase (Supplementary Fig. 13f), and activated a NF-κB reporter construct alone or in combination with IL-1 (Fig. 3f), which was inhibited by coexpression of K7 (Fig. 3g). Expression of WNK3 stimulated IL-8 production alone or in combination with IL-1 (Supplementary Fig. 13g). Short interfering RNA (siRNA)-mediated knockdown of various WNK family members resulted in increased growth of VSV (Fig. 3h and Supplementary Fig. 13h). These results illustrate the value of our proteomics data set by revealing a previously unknown role for WNK kinases in the antiviral immune response.

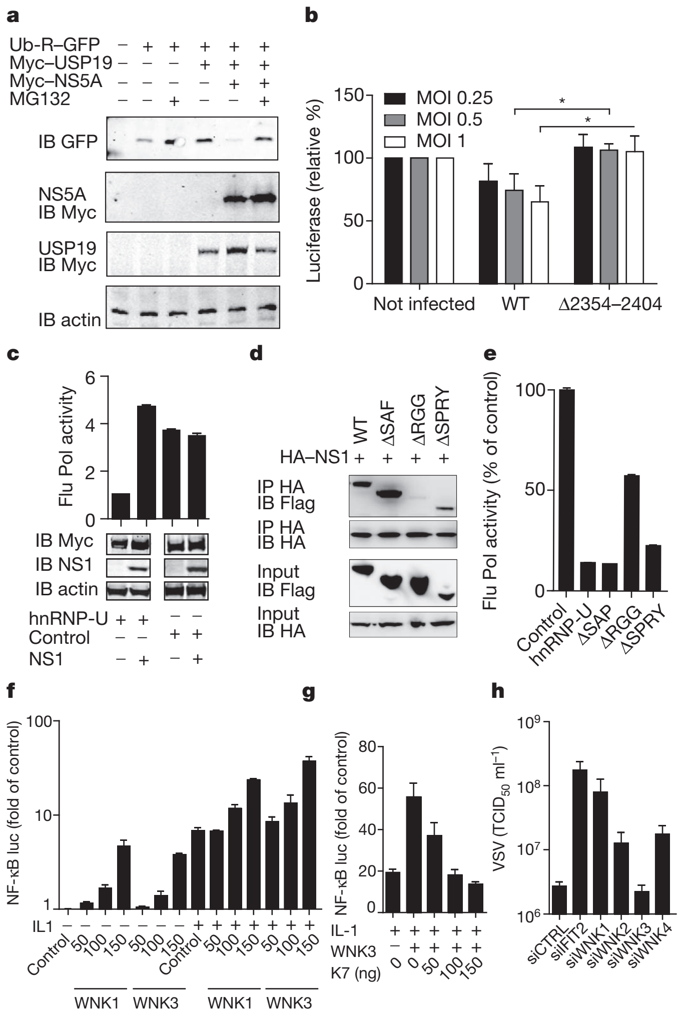


Figure 3 | Functional validation of USP19, hnRNP-U and WNK kinases as viral targets. **a**, 293T cells transfected with GFP fused to a proteasomal degradation signal (Ub-R–GFP), Myc-tagged USP19 and NS5A. MG132 was added for 6 h, and cells were analysed by immunoblotting (IB). **b**, Huh7.5 cells expressing firefly luciferase were infected with wild-type (WT) HCV or HCV lacking the USP19 interaction site (Δ2354–2404). Results are activities after 96 h (means ± s.d. for three independent experiments). Asterisk, $P < 0.025$, Student's *t*-test. MOI, multiplicity of infection. **c**, FluAV minireplicon activity in the presence of 33 ng of Myc–hnRNP-U or GRB2 (control) and 33 ng of NS1 (A/PR/8/34). Results are FluAV polymerase (Pol) activity (means ± s.d. for duplicate measurements, one representative of three). Immunoblots show protein expression (24 h). **d**, Co-immunoprecipitation of Flag–hnRNP-U and indicated mutants with HA–NS1 (A/PR/8/34). **e**, As in **c**, but for 100 ng of GFP (control) and Flag–hnRNP-U mutants (means ± s.d. for duplicate measurements, one representative of three). **f**, **g**, NFκB-luciferase (luc) activity in HEK293 cells in the presence of increasing amounts (in ng) of WNK1 and WNK3 with or without IL-1 (**f**), and K7 (**g**). Results are measurements after 24 h (means ± s.d. for triplicate experiments). **h**, HeLa cells transfected with siRNAs against WNKs and non-silencing control were infected. $TCID_{50}$, 50% tissue culture infective dose (mean ± s.d., $n = 3$).

Proteomic profiling of such a large group of viral regulators of cell function offers the opportunity to explore kinship in their mode of action and, by inference, the perturbation strategy of the viruses that encode them. We defined a notion of kinship distance by incorporating shared targets, proximity in the humPPI of non-shared targets, and their strength of interactions. viORFs from the same viral family had

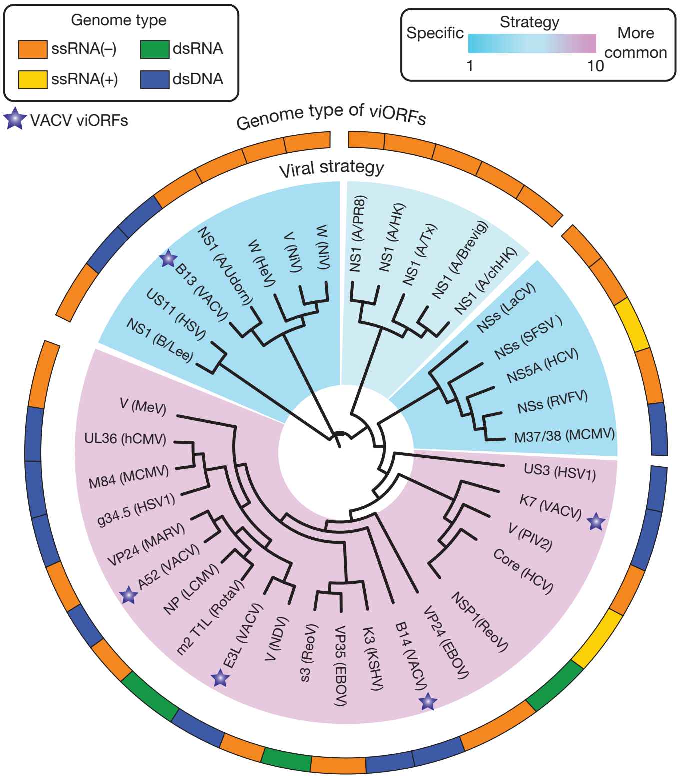


Figure 4 | Similarities of viORF actions. Dendrogram of viORF relationships based on the kinship distance, which integrates the number of shared targets and the network distance in the humPPI of the distinct targets. The virus genotype that the individual viORF derives from is shown in a colour code in the circle around the dendrogram. EBOV, Ebola virus; hCMV, human cytomegalovirus; HCV, hepatitis C virus; HeV, Hendra virus; HSV, herpes simplex virus; HSV1, herpes simplex virus 1; KSHV, Kaposi's sarcoma-associated herpesvirus; LaCV, La Crosse virus; LCMV, lymphochoriomeningitis virus; MARV, Marburg virus; MCMV, murine cytomegalovirus; MeV, measles virus; NDV, Newcastle disease virus; NiV, Nipah virus; PIV2, parainfluenza virus 2; ReoV, reovirus; RotaV, rotavirus; SFSV, sandfly fever sicilian virus. viORFs from VACV are indicated with a star.

short average kinship distances (Supplementary Fig. 14), consistent with their evolutionary relationship. Notable exceptions were viORFs from paramyxoviruses, which had an average distance even larger than randomized viral target profiles, possibly reflecting a particularly pleiotropic mechanism of action. We generated a dendrogram that showed the kinship distance of the individual viORFs as a proxy for the perturbation strategy of the cognate virus (Fig. 4). Roughly half of the viORFs clustered in a central, rather dense part of the tree, reflecting overlapping strategies, whereas the other half was more distant, probably indicating more unique targeting strategies. Many clusters represented viORFs from evolutionarily related viruses, which are more likely to exercise comparable perturbation strategies. For example, most influenza A virus NS1 proteins and all NSs proteins from bunyaviruses clustered together. A few viORFs did not cluster according to their genome group, which was evocative of some degree of evolutionary convergence with the proteins of other viruses on shared pathways, or more distinctive routes of action, possibly as part of a combined attack with another ORF of the same virus. This is best illustrated by the five viORFs from VACV, which were found scattered in the tree and were likely to have evolved to fulfil specific, complementary functions.

Our results demonstrate that viruses have evolved to exploit a variety of cellular mechanisms, and suggest that the host cell relies on the proper homeostatic regulation across these diverse cellular processes to detect, alert to and counteract pathogen interference. In addition, the study provides a rationale for considering or excluding the targeting of specific intracellular pathways for pan-viral or virus-specific antiviral therapy.

METHODS SUMMARY

Complementary DNA of tandem affinity-tagged viORFs was amplified by polymerase chain reaction and cloned into the pTO-SII-HA-GW vector by using Gateway recombination (Invitrogen). The resulting plasmids were used to generate hygromycin-selected stable isogenic HEK293 Flp-In TREx cell lines, and viORF expression was stimulated by doxycycline[12]. Protein complexes isolated by tandem affinity purification using Strep-II and haemagglutinin (HA)-affinity reagents were analysed by LC–MS/MS with an LTQ Orbitrap XL, an LTQ Orbitrap Velos or a QTOF mass spectrometer. The data were searched against the human SwissProt protein database, using Phenyx and Mascot. The humPPI was generated using public interaction databases. Recombinant HCVs (strain JC1) with mutations in domain III of NS5A were generated by transfecting full-length genomic RNA with targeted deletions in the NS5A region. Subcellular localization of proteins was performed on a Leica SP2 confocal microscope. The influenza virus replicon assay was performed as described previously[12].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 September 2011; accepted 7 June 2012.
Published online 18 July 2012.

1. Watanabe, T., Watanabe, S. & Kawaoka, Y. Cellular networks involved in the influenza virus life cycle. *Cell Host Microbe* **7,** 427–439 (2010).
2. Goff, S. P. Host factors exploited by retroviruses. *Nature Rev. Microbiol.* **5,** 253–263 (2007).
3. Finlay, B. B. & McFadden, G. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* **124,** 767–782 (2006).
4. Bowie, A. G. & Unterholzner, L. Viral evasion and subversion of pattern-recognition receptor signalling. *Nature Rev.* **8,** 911–922 (2008).
5. Jager, S. *et al.* Global landscape of HIV–human protein complexes. *Nature* **481,** 365–370 (2012).
6. Shapira, S. D. *et al.* A physical and regulatory map of host–influenza interactions reveals pathways in H1N1 infection. *Cell* **139,** 1255–1267 (2009).
7. Uetz, P. *et al.* Herpesviral protein networks and their interaction with the human proteome. *Science* **311,** 239–242 (2006).
8. Krishnan, M. N. *et al.* RNA interference screen for human genes associated with West Nile virus infection. *Nature* **455,** 242–245 (2008).
9. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319,** 921–926 (2008).
10. Konig, R. *et al.* Human host factors required for influenza virus replication. *Nature* **463,** 813–817 (2010).
11. Karlas, A. *et al.* Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **463,** 818–822 (2010).
12. Pichlmair, A. *et al.* IFIT1 is an antiviral protein that recognizes 5′-triphosphate RNA. *Nature Immunol.* **12,** 624–630 (2011).
13. Li, S., Wang, L., Berman, M., Kong, Y. Y. & Dorf, M. E. Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. *Immunity* **35,** 426–440 (2011).
14. Cravatt, B. F., Simon, G. M. & Yates, J. R. III. The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991–1000 (2007).
15. Randall, R. E. & Goodbourn, S. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J. Gen. Virol.* **89,** 1–47 (2008).
16. Schmolke, M. & Garcia-Sastre, A. Evasion of innate and adaptive immune responses by influenza A virus. *Cell. Microbiol.* **12,** 873–880 (2010).
17. Rodriguez, J. J., Wang, L. F. & Horvath, C. M. Hendra virus V protein inhibits interferon signaling by preventing STAT1 and STAT2 nuclear accumulation. *J. Virol.* **77,** 11842–11845 (2003).
18. Glatter, T., Wepf, A., Aebersold, R. & Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol. Syst. Biol.* **5,** 237 (2009).
19. Lynn, D. J. *et al.* InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.* **4,** 218 (2008).
20. Burkard, T. R. *et al.* Initial characterization of the human central proteome. *BMC Syst. Biol.* **5,** 17 (2011).
21. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
22. Hassink, G. C. *et al.* The ER-resident ubiquitin-specific protease 19 participates in the UPR and rescues ERAD substrates. *EMBO Rep.* **10,** 755–761 (2009).
23. Arima, N. *et al.* Modulation of cell growth by the hepatitis C virus nonstructural protein NS5A. *J. Biol. Chem.* **276,** 12675–12684 (2001).
24. Lu, Y. *et al.* USP19 deubiquitinating enzyme supports cell proliferation by stabilizing KPC1, a ubiquitin ligase for $p27^{Kip1}$. *Mol. Cell. Biol.* **29,** 547–558 (2009).
25. Valente, S. T. & Goff, S. P. Inhibition of HIV-1 gene expression by a fragment of hnRNP U. *Mol. Cell* **23,** 597–605 (2006).
26. Hasegawa, Y. *et al.* The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* **19,** 469–476 (2010).

27. Schroder, M., Baran, M. & Bowie, A. G. Viral targeting of DEAD box protein 3 reveals its role in TBK1/IKKepsilon-mediated IRF activation. *EMBO J.* **27,** 2147–2157 (2008).
28. Moniz, S. & Jordan, P. Emerging roles for WNK kinases in cancer. *Cell. Mol. Life Sci.* **67,** 1265–1276 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Basler, A. Bergthaler, K.-K. Conzelmann, A. Garcia-Sastre, M. Hardy, W. Kaiser, E. Mühlberger, R. Randall, B. Roizman, N. Ruggli, B. Sherry and T. Wolf for providing viral ORF cDNAs; P. Jordan for providing WNK expression constructs; S. Nakagawa for Flag-hnRNP-U; M. Sophie-Hiet for GFP-NS5A; M. Zayas for the pFK-Jc1-NS5A-HA expression plasmid; E. Rudashevskaya, A. Stukalov, F. Breitwieser and M. Trippler for support; H. Pickersgill and T. Brummelkamp for critically reading the manuscript; C. Baumann for discussions; and M. Vidal for discussions and for sharing unpublished information. The work was funded by the Austrian Academy of Sciences, an i-FIVE European Research Council grant to G.S.-F., a European Molecular Biology Organization long-term fellowship to A.P. (ATLF 463-2008), Science Foundation Ireland grant 07/IN1/B934 to O.M. and A.G.B., Deutsche Forschungsgemeinschaft grants We 2616/5-2 and SFB 593/B13 to F.W., Ko1579/5-1 to G.K., and FOR1202, TP1 to R.B., and the German Ministry for Education and Research (Suszeptibilität bei Infektionen: HCV; TP1, 01KI 0786) to R.B. J.C. is funded by the Austrian Ministry of Science and Research (GEN-AU/BIN).

Author Contributions A.P., G.A., O.M., R.S., M.H., M.B., A.S., C.A.E., A.G., A.C.M., A.F., C.H., S.G., F.W. and G.K. performed experiments. A.P. and G.S.-F. conceived the study. A.P., G.A., R.B., A.G.B. and G.S.-F. designed experiments. K.K., A.C.M., K.L.B. and J.C. performed mass spectrometry and bioinformatic data analysis. T.B., K.L., S.G., G.K., F.W., R.B. and A.G.B. provided critical material. All authors contributed to the discussion of results and participated in manuscript preparation. A.P., K.K., J.C. and G.S.-F. wrote the manuscript.

Author Information The protein interactions from this publication have been submitted to the IMEx consortium (http://imex.sf.net) through IntAct (identifier IM-17331). Mass spectrometry data are available at http://inhibitomev1.sf.net; microarray data were deposited in ArrayExpress (accession number E-MTAB-1148). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.S.-F. (gsuperti@cemm.oeaw.ac.at).

METHODS

Plasmids, viruses and reagents. Expression constructs were generated by PCR amplification of viORFs followed by Gateway cloning (Invitrogen) into the plasmids pCS2-6myc-GW, pCMV-HA-GW and pTO-SII-HA-GW. pCAGS-Flag-hnRNP-U and mutants thereof were provided by S. Nakagawa. Ub-R–GFP and Myc–USP19 were published previously[22]. pHA-PIK3R2 was from Oliver Hantschel. GFP–NS5A domain mutants were published previously[30]. Recombinant HCV variants with mutations in domain III of NS5A were generated by replacing the NS5A fragment in pFK-Jc1-NS5A-HA, containing the full-length HCV chimaeric Jc1 genome[31] in which a HA tag is inserted in frame within NS5A and in pFK-JcR-2a containing *Renilla* luciferase fused amino-terminally with the 16 N-terminal amino-acid residues of the core protein and C-terminally with the foot-and-mouth disease 2A peptide coding region, enabling direct quantification of viral replication by measuring *Renilla* luciferase activity[32]. All viruses were produced by transient transfection of Huh7.5 cells with RNA transcribed *in vitro*. Recombinant RVFV (Rift valley fever virus)[33] expressing tandem affinity-tagged (GS-TAG) versions of NSs proteins were generated by replacing the RVFV NSs open reading frame with GS-tagged versions of NSs that were generated by PCR amplification. The FluAV minireplicon system to measure FluAV polymerase activity[34], IFN-β–luciferase, NF-κB-luciferase and the *Renilla* luciferase control plasmid (pRL-TK; Promega) were described previously[35].

Streptavidin beads were from IBA (Strep-Tactin agarose); HA–agarose (clone HA7) was from Sigma. Antibody against β-tubulin (anti-β-tubulin; clone DM1A) was from Abcam, anti-β-actin (catalogue number AAN01) was from Cytosceleton. IRDye-conjugated anti-c-Myc (catalogue number 600-432-381) and anti-rabbit (catalogue number 611-732-127) secondary reagents were from Rockland. Alexa Fluor 680-conjugated goat anti-mouse (catalogue number 10524963) were from Molecular Probes. Reagents for quantitative RT–PCR were from Qiagen. Poly(dA)•poly(dT) were from Sigma and transfected with Lipofectamine 2000 (Invitrogen) or Polyfect (Qiagen). Stimulatory PPP-RNA was described previously[12]. MG132 was from Sigma. IFN-β and IFN-α2a were from PBL Interferonsource. Tumour necrosis factor-α and IL-1β were from Pierce. IL-8 was measured by enzyme-linked immunosorbent assay (BD). Lymphochoriomeningitis virus (Armstrong strain), FluAV (A/PR/8/34), VSV (Indiana strain) and VSV-M2 (mutant VSV with M51R substitution of the matrix protein, leading to IFN-α/β induction; originally called AV3) have been described previously[12]. Virus titres were measured by determining the half-maximal infectious dose ($TCID_{50}$) on Vero cells, or on Huh7.5 cells for HCV.

Cells, co-immunoprecipitations and imaging. HEK293 Flp-In TREx cells that allow doxycycline-dependent transgene expression were from Invitrogen. HEK293, 293T, HeLa S3 (ref. 12), Lunet, Lunet-Neo-sgNS5A(RFP), Huh7/5.2 and Huh7.5 cells have been described previously[30]. Highly permissive Huh7.5 or Huh7.5 FLuc, stably expressing firefly luciferase introduced by lentiviral transduction[32], were used for HCV infection experiments. Fibroblasts were kept in DMEM medium (PAA Laboratories) supplemented with 10% (v/v) FCS (Invitrogen) and antibiotics (100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin). For inducible transgene expression, HEK293 Flp-In TREx cells were treated for 24–48 h with doxycycline (1 μg ml^{-1}), depending on cellular density to just about reach confluence. For siRNA-mediated knockdown, if not stated otherwise in figure legends, 5 nmol of siRNA pool (Supplementary Table 7) was mixed with HiPerfect (Qiagen) and added to 10^5 HeLa S3 cells. After 48 h, cells were used for experiments. For co-immunoprecipitations 293T cells were transfected with expression plasmids for 24–48 h and lysates were used for affinity purification in TAP buffer[12] using anti-HA–agarose or anti-c-Myc-coated beads. For protein detection in western blot analysis a Li-Cor infrared imager was used. Confocal images were acquired with a Leica SP2 confocal microscope.

Affinity purification, mass spectrometry and transcriptome analysis. HEK293 Flp-In TREx cells and isolation of protein complexes by TAP and peptide analysis by LC–MS/MS have been described previously[18]. Proteins identified by this method can be found in a complex but do not necessarily bind directly to each other. In brief, five subconfluent 15-cm dishes of cells were stimulated with 1 μg ml^{-1} doxycycline for 24–48 h. Protein complexes were isolated by TAP using streptavidin agarose followed by elution with biotin, and a second purification step using HA–agarose beads. Proteins were eluted with 100 mM formic acid, neutralized with triethylammonium bicarbonate (TEAB) and digested with trypsin, and the peptides were analysed by LC–MS/MS [36]. For bunyavirus NSs proteins, recombinant viruses[33] containing GS-tagged NSs proteins were generated. Protein complexes were denatured in Laemmli buffer[37] and separated by one-dimensional SDS–PAGE; entire lanes were excised and digested *in situ* with trypsin and the resultant peptides were analysed by LC–MS/MS. Mass spectrometric analysis was performed for gel-free and gel-based samples, respectively, on a hybrid LTQ Orbitrap XL, an LTQ Orbitrap Velos mass spectrometer (both from ThermoFisher Scientific) or on a quadrupole time-of-flight mass spectrometer (QTOF Premier; Waters) coupled to an 1100/1200 series high-performance liquid chromatography system (Agilent Technologies). Data generated by LC–MS/MS were searched against the human SwissProt protein database (v. 2010.09, plus appended viral bait proteins) with Mascot (v. 2.3.02) and Phenyx (v. 2.6). One missed tryptic cleavage site was allowed. Carbamidomethyl cysteine was set as a fixed modification, and oxidized methionine was set as a variable modification. A false-positive detection rate of less than 1% on the protein groups was imposed (Phenyx z-score more than 4.75 for single peptide identifications, z-score more than 4.2 for multiple peptide identifications; Mascot single peptide identifications ion score more than 40, multiple peptide identifications ion score more than 14).

To measure gene expression, Huh7/5-2 cells were left uninfected or infected with HCV (strain JC1) at a MOI of 5, and RNA was isolated using Trizol (Invitrogen) after 4, 12, 24, 48 and 72 h. Gene expression analysis was performed in duplicate using an Affymetrix platform (Affymetrix Human Genome U133A 2.0 Array).

Bioinformatic analysis. Data filtering. All proteins identified in the GFP negative controls (51 proteins) were removed.

Data normalization. Affinity-purification MS experiments were performed with two biological replicates and two technical replicates for each; that is, four replicates. We first normalized individual replicates according to the NSAF procedure[29]. The replicates of each viORF normalized data element were then assembled in a table with 0 for missing detection, and each viral target was assigned the average NSAF value across the replicates. On the basis of a robust estimate (MAD) of the coefficient of variation (Supplementary Fig. 15a) we further penalized highly variable targets by applying a reduction factor between 1 (modest variability) and 0.5 (high variability) (Supplementary Fig. 15b). Direct normalization through a division by the standard deviation was excluded because of the limited number of replicates available. For a given viORF v and a viral target p, the weight given to the interaction v–p was hence computed as

$$\text{strength}_{v,p} = \text{mean}(\text{NSAF}_{v,p,i})\text{reduction}[\text{CV}(\text{NSAF}_{v,p,i})]$$

where i accounts for the replicates. The distribution of strength values is shown in Supplementary Fig. 15c.

Human interactome. We integrated human physical protein–protein interactions (humPPI) obtained from public databases (IntAct, BioGRID, MINT, HPRD and InnateDB[19]) and thereby obtained an interactome (largest connected component) comprising 13,350 proteins and 90,292 interactions.

Human central proteome. A list of commonly expressed human proteins was assembled by merging a previous study[20] with mapped (orthologues) mouse proteins found in the intersection of six mouse tissues[38] and genes expressed in all except four or fewer tissues from SymAtlas. The resulting list included 4,276 proteins and is provided as Supplementary Table 8.

Network topological measures. We retained two classical measures: the connectivity (degree)—that is, the number of interactions of one protein in the PPI—and the relative betweenness centrality, which is equal to the relative number of shortest paths between any two proteins that go through a given protein.

MS-weighted measures. To compute a weighted characteristic of the targeted host proteins, for example connectivity in the human PPI, of one viral modulator vm we used

$$\text{weighted\_connectivity}(vm) = \sum\nolimits_{p \in \mathrm{T}(vm)} \alpha_p \text{connectivity}(p)$$

where $\mathrm{T}(vm)$ is the set of all human proteins targeted by vm; α_p were proportional to the estimated interaction strength, and sum to 1. When the same viral modulator was considered in several viruses (for example NS1 of FluAV), we computed the weights for each interacting protein taking the maximum of the strengths found in different viruses to avoid any bias by over-represented viral modulators; that is, $\alpha_p \propto \max_{v \in \text{NS1\_viruses}} \text{strength}_{v,p}$. Null distributions were generated by assigning actual weights to random proteins 10,000 times, thereby obtaining a histogram of 10,000 random weighted characteristics, which was fitted with a gamma distribution to estimate P values (Supplementary Fig. 15d).

Weighted functional annotation analysis. We performed GO and KEGG pathways analysis integrating the interaction strengths of viORF targets by summing all the above normalized (sum equal to 1) α_p weights found in a GO term or a pathway to obtain a score. This score was then compared with a null distribution modelled by a gamma fit on 1,000 random scores to estimate a P value. Random scores were obtained by assigning the weights to random proteins and summing those that fell in the GO term or pathway.

Perturbation map and relative position along a pathway. These two computations were performed in accordance with published methods[20]. Pathways were taken from NCI-PID[39], and the perturbation map algorithm (GO fluxes in ref. 20) was modified to use the interaction strengths between viORFs and their targets as weights in scoring interaction between GO terms instead of constant weights. For simplification, GO terms were reduced to 14 categories (Supplementary Table 9).

Perturbation map null distributions were obtained with 250 randomized annotated networks that respected the original network connectivity distribution and GO term frequencies.

Distance of viORFs. Given two viORFs x and y, the distance $d(x,y)$ is defined as follows. Let S be the union of all x and y targets, D_x the targets unique to x, and D_y those unique to y. A preliminary distance c is computed by summing all the human interactome shortest path distances from individual targets in D_x and D_y with the targets unique to the other viORF, considering interaction strengths to penalize differences on strong different targets and minimize the impact of weaker distinct targets. Thus,

$$c = \sum_{a \in D_x} \text{strength}_{x,a} \times \text{shortest}(a,D_y) + \sum_{a \in D_y} \text{strength}_{y,b} \times \text{shortest}(b,D_x)$$

Finally, c is normalized to take into account the number of distinct targets compared with the total number of targets: $d(x,y) = c(|D_x \cup D_y|)/|s|$, where $|...|$ denotes set cardinality—that is, the number of elements.

The random distance distributions were obtained as follows: for each viORF, its targets were replaced by a random selection of the same number of proteins from the humPPI such that the same pairs of (random) distances could be computed. The overall procedure was repeated 100 times and in the case of the HEK293 selection the human proteins randomly chosen were restricted to the humPPI and to proteins identified by mass spectrometric analysis of the HEK293 proteome[20].

29. Zybailov, B. *et al.* Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5,** 2339–2347 (2006).
30. Appel, N. *et al.* Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog.* **4,** e1000035 (2008).
31. Pietschmann, T. *et al.* Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc. Natl Acad. Sci. USA* **103,** 7408–7413 (2006).
32. Reiss, S. *et al.* Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell Host Microbe* **9,** 32–45 (2011).
33. Habjan, M., Penski, N., Spiegel, M. & Weber, F. T7 RNA polymerase-dependent and -independent systems for cDNA-based rescue of Rift Valley fever virus. *J. Gen. Virol.* **89,** 2157–2166 (2008).
34. Dittmann, J. *et al.* Influenza A virus strains differ in sensitivity to the antiviral action of Mx-GTPase. *J. Virol.* **82,** 3624–3631 (2008).
35. Keating, S. E., Maloney, G. M., Moran, E. M. & Bowie, A. G. IRAK-2 participates in multiple Toll-like receptor signaling pathways to NFκB via activation of TRAF6 ubiquitination. *J. Biol. Chem.* **282,** 33435–33443 (2007).
36. Haura, E. B. *et al.* Using iTRAQ combined with tandem affinity purification to enhance low-abundance proteins associated with somatically mutated EGFR core complexes in lung cancer. *J. Proteome Res.* **10,** 182–190 (2010).
37. Burckstummer, T. *et al.* An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nature Methods* **3,** 1013–1019 (2006).
38. Kislinger, T. *et al.* Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125,** 173–186 (2006).
39. Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37,** 674–679 (2009).

doi:10.1038/nature11218

Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration

Panteleimon Rompolas[1], Elizabeth R. Deschene[1]*, Giovanni Zito[1]*, David G. Gonzalez[2], Ichiko Saotome[1], Ann M. Haberman[2] & Valentina Greco[1]

Tissue development and regeneration depend on cell–cell interactions and signals that target stem cells and their immediate progeny[1]. However, the cellular behaviours that lead to a properly regenerated tissue are not well understood. Using a new, non-invasive, intravital two-photon imaging approach we study physiological hair-follicle regeneration over time in live mice. By these means we have monitored the behaviour of epithelial stem cells and their progeny[2–4] during physiological hair regeneration and addressed how the mesenchyme[5] influences their behaviour. Consistent with earlier studies[6], stem cells are quiescent during the initial stages of hair regeneration, whereas the progeny are more actively dividing. Moreover, stem cell progeny divisions are spatially organized within follicles. In addition to cell divisions, coordinated cell movements of the progeny allow the rapid expansion of the hair follicle. Finally, we show the requirement of the mesenchyme for hair regeneration through targeted cell ablation and long-term tracking of live hair follicles. Thus, we have established an *in vivo* approach that has led to the direct observation of cellular mechanisms of growth regulation within the hair follicle and that has enabled us to precisely investigate functional requirements of hair-follicle components during the process of physiological regeneration.

Although stem cells and their immediate progeny are critical for tissue regeneration, we still lack knowledge concerning the discrete sequential steps that lead to proper tissue regeneration. Available methods to address these questions during regeneration are largely static and provide only snapshots of this highly dynamic process. An alternative approach is to continuously visualize stem cells and their progeny throughout physiological regeneration. Recent technological advances have enabled stem cell imaging in mammalian tissues *in vivo*, such as haematopoietic stem cells in the bone marrow[7,8]. However, these studies are inherently invasive and therefore it is not possible to monitor the same stem cells for long periods (days to weeks) using these methods.

To study a physiological tissue regeneration process, we established a method to image entire skin hair follicles over time in live mice. The skin is a suitable tissue for live imaging owing to its accessibility[9,10]. In addition, hair follicles continuously regenerate through alternating rest (telogen) and growth (anagen) phases (Supplementary Fig. 1), enabling us to study stem cell behaviours during physiological regeneration. Furthermore, hair follicles are anatomically well defined and the location of different cell types is well characterized[2,6,11,12]. During the quiescent phase, two populations coexist in the lower portion of the hair follicle; the stem cells, which are localized in a specialized region called the bulge, and their immediate progeny within the hair germ (Fig. 1a, b). The stem cell progeny is in contact with a mesenchymal group of cells—called dermal papilla (Fig. 1a)—that is able to induce hair regeneration (Supplementary Fig. 1) [5]. Our recent work has shown that signals released by the mesenchyme activate the progeny at the beginning of the hair regeneration cycle[6]. Thus, hair regeneration initially relies mainly on the expansion of the progeny and not of the stem cells. Subsequently, stem cells in the bulge will slowly contribute to hair regeneration over time[6,12–14].

To visualize directly the hair-follicle stem cells and progeny, we used a transgenic mouse line that marks all of the epithelial nuclei in the skin by expressing a fusion protein of histone H2B with green fluorescent protein (GFP) driven by the keratin 14 promoter (K14H2BGFP)[3,6]. This transgenic mouse provides a strong nuclear signal that can resolve individual cells. Using a two-photon laser scanning microscope we acquired serial optical sections from the skin of anaesthetized 3-week-old mice (Supplementary Fig. 2 and Supplementary movie 1), which enabled us to visualize the entire hair follicle and identify the location of the stem cells and their progeny based on their distinct morphological features (Fig. 1b and Supplementary Fig. 3).

During the early stages of regeneration, proliferation is first initiated in the stem cell progeny, which undergo a rapid downward expansion (Supplementary Fig. 1). We asked whether this asymmetrically organized growth of the hair follicle relies on spatial regulation of processes such as cell divisions. To test this idea, we performed time-resolved imaging of hair follicles at the beginning of the hair regeneration cycle (around postnatal day 22 (P22)) to visualize cell proliferation over several hours. By these means, we have recorded several mitotic events with clear spatial and temporal resolution (Fig. 1c and Supplementary movie 2). Systematic quantification of the location and axis of divisions from several recordings shows that most divisions occur in the lower part of the follicle at the interface between the progeny and the mesenchyme at the beginning of a new growth (Fig. 1d). At this stage, the axis of division is parallel to the long axis of hair follicle growth. A second, less pronounced, centre of growth was identified in the upper part of the progeny (Fig. 1d). Cell divisions were observed within the bulge even at lower frequencies, (Fig. 1d, e, and Supplementary movies 3 and 4). The axes of these bulge divisions varied, ranging from perpendicular to parallel with respect to the long axis of hair follicle growth (Fig. 1d, e). Hair growth in the mouse is a fairly synchronized process at 3 weeks of age. To test whether this synchronization is the result of temporal regulation of cell divisions within the hair follicle, we quantified the time when cell divisions occurred within a single hair follicle and in relation to neighbouring hair follicles (at approximately P22). In some of our recordings, clustering was observed between cell divisions within one follicle as well as between adjacent follicles. However, these behaviours were not consistent in every recording (Supplementary Fig. 4 and Supplementary movie 2).

During regeneration, the lower part of the hair follicle undergoes a major reorganization as the mesenchyme transitions from being extruded from the epithelium to being encompassed by it[15] (Supplementary Fig. 1). This led us to propose that changes in tissue architecture may rely on other cellular processes in addition to cell divisions. Analysing several of our recordings we observed that some follicles rapidly stretch downwards during this growth phase (at approximately P23; Fig. 2a,

[1]Department of Genetics, Department of Dermatology, Yale Stem Cell Center, Yale Cancer Center, Yale School of Medicine, New Haven, Connecticut 06510, USA. [2]Department of Laboratory Medicine, Yale School of Medicine, New Haven, Connecticut 06510, USA.
*These authors contributed equally to this work.

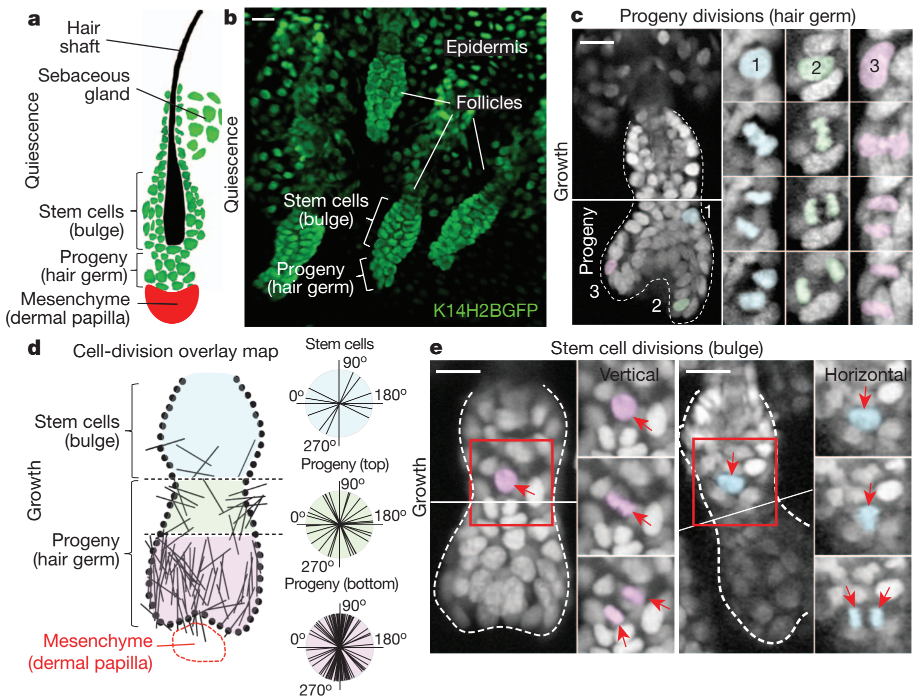


Figure 1 | Cell divisions are spatially regulated in the hair follicle at the beginning of a new growth. **a**, A hair follicle in quiescence. Distinct populations of cells that participate in hair regeneration, including stem cells, progeny and mesenchyme, reside in defined anatomical compartments of the hair follicle. **b**, A three-dimensional reconstruction of quiescent live hair follicles from serial optical sections acquired by two-photon laser scanning microscopy. Epithelial nuclei (green) are made visible using the H2B–GFP fusion protein driven by the keratin 14 promoter (K14H2BGFP). **c**, An example of progeny cell division. A single optical section of a live hair follicle (left panel) and magnified views (right, insets) of three nuclei in the progeny compartment that undergo mitosis (right). **d**, The locations and axes of cell divisions were quantified from several hair follicles ($n = 17$) at early growth phase (anagen II). **e**, Two examples of vertical (left panel) and horizontal (right panel) stem cell divisions. A single optical slice of live hair follicles (left, insets) and magnified views of nuclei in the stem cell compartment (right, insets) undergoing mitosis. Red arrows, parental and daughter nuclei during mitosis. Time lapse between sequential frames is about 15 min (**c**) and 45 min (**e**). Scale bars, 20 μm.

Supplementary Fig. 5, and Supplementary movies 5 and 6). To test whether this tissue extension is the result of a morphological rearrangement of the tissue, we quantified the distance between adjacent nuclei within the stem cell or the progeny compartment. The distance between nuclei increased specifically within the progeny, which may account for the overall downwards extension of the hair follicle (Fig. 2a and Supplementary Fig. 5). This type of behaviour is reminiscent of morphogenetic movements observed during development in classical model organisms such as *Drosophila* or zebrafish (*Danio*) embryos[16] and provides a previously unidentified mechanism of hair growth in addition to cell division.

To dissect further these novel dynamic processes, we analysed time-resolved optical sections at the coronal plane of the follicles where the epithelium interfaces the mesenchyme. During growth, *in vivo* recordings revealed a major reorganization of the epithelial stem cell progeny surrounding the mesenchymal dermal papilla. Over 4 h the nuclei transition from a disorganized pattern to a single row aligned around the mesenchyme (Fig. 2b and Supplementary movie 7). In addition, the lower epithelial part of the follicle constricts as it encompasses the mesenchyme (Fig. 2b; 0 h versus 4 h). This major epithelial nuclear reorganization occurs concurrently in adjacent follicles (Supplementary Fig. 6 and Supplementary movie 7). Furthermore, in more advanced growth stages, long-range migrations within the outer most layer (outer root sheath) of the lower hair follicle were observed (Fig. 2c and Supplementary movie 8). Recent data using lineage-tracing approaches have indicated that stem cells can migrate out of the bulge, either downwards towards the progeny or upwards towards the sebaceous gland (Fig. 1a)[17,18]. We did not observe downwards

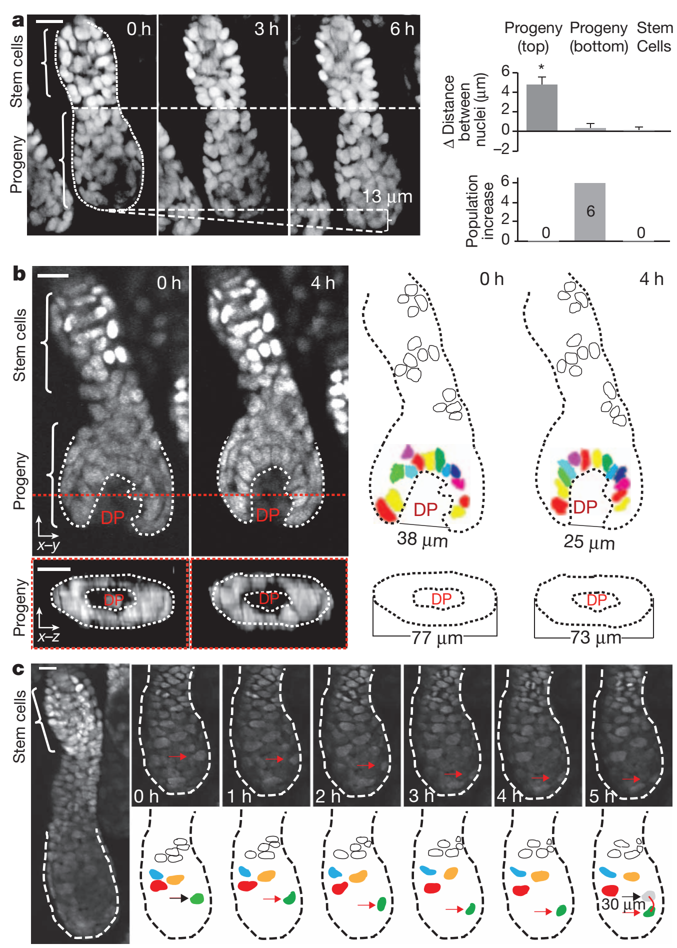


Figure 2 | The stem cell progeny compartment undergoes morphological reorganization during growth. **a**, Downwards extension of follicles during growth. Optical sections at three consecutive time points (3 h apart) of a growing live hair follicle shows the progeny compartment extending downwards (left three panels). The increase in the distances between nuclei, and total cell numbers were quantified within the stem cells and the progeny compartments (approximately anagen II to IIIa) (right panel, data are expressed as mean ± s.e.m., $n = 13$–20; asterisk, $P < 0.0001$; see also Supplementary Fig. 5). **b**, Reorganization of nuclei within the hair follicle. Two optical sections (left panels), corresponding traces and measurements (right panels) at the coronal and transverse planes (*xy* and *xz*) of the same follicle at time 0 and 4 h, respectively (approximately anagen II to IIIa) (bottom panels). **c**, Downwards migration of nuclei in the hair follicle during growth. Single optical sections showing complete (left panel) and lower partial views (top panels) of a single hair follicle at successive time points, 1 h apart. Red arrows in the optical slices and corresponding traces (bottom panels) mark a nucleus that is moving downwards, covering a distance of ~30 μm within 5 h (approximately anagen IIIb). The position of the green nucleus at 0 h is shown in grey for comparison (bottom right panel). DP, dermal papilla. Scale bars, 20 μm.

migrations of the stem cells to the progeny but captured a short upwards migration within the bulge stem cells (Supplementary Fig. 7 and Supplementary movie 9). Based on these data, we suggest that migratory events within the bulge may be temporally regulated or may take place at a much slower pace than we can resolve in the timeframe of our experiments (3–14 h). Taken together, our findings reveal new dynamic cellular processes adopted by the stem cells and their immediate progeny during physiological regeneration that would have been missed by conventional static analysis.

Epithelial–mesenchymal interactions are crucial for the development and regeneration of many organs such as limbs[19]. In the hair follicle, the mesenchymal dermal papilla is a key signalling centre, able to induce *de novo* hair-follicle formation after transplantation[5]. Moreover, previous work has identified mesenchymal signals, including signalling by FGF7 and FGF10, and BMP inhibitors, as regulators for the initiation of the hair regeneration cycle[6,20]. These and other data suggest that the mesenchyme is sufficient to induce hair regeneration. However, the requirement of the mesenchyme for initiation of hair regeneration has not been tested. To be able to selectively eliminate the mesenchyme, we set up a laser-induced cell-ablation approach to target fluorescently labelled dermal papilla cells using a Lef1RFP transgenic mouse (expressing red fluorescent protein under the control of a Lef1 promoter fragment)[6,21] at the beginning of a new hair growth (at approximately P19; Fig. 3a and Supplementary Fig. 8). Analysis of the tissue immediately after laser ablation showed that the dermal papilla was disrupted, whereas adjacent tissues, such as the progeny or the overlying epidermis, remained undamaged (Supplementary Fig. 8). To assess the long-term effects of dermal papilla ablation on hair regeneration, we revisited the same hair follicles in separate imaging sessions, hours or even days apart (Fig. 3a). When the same hair follicles were observed at 4 and 7 days after ablation, follicles with ablated mesenchymal dermal papillae appeared quiescent (Fig. 3b). In contrast, surrounding follicles with intact mesenchyme grew at a normal pace (Fig. 3c, e). To determine whether the mesenchymal ablation altered the molecular and cellular composition of the hair follicle, skin samples containing mesenchymal ablated follicles were collected and analysed at 7 days post ablation. Histological and immunofluorescence data showed that the stem cells and their progeny maintain their molecular properties after dermal papilla ablation (Supplementary Fig. 9). Furthermore, the ablated follicles

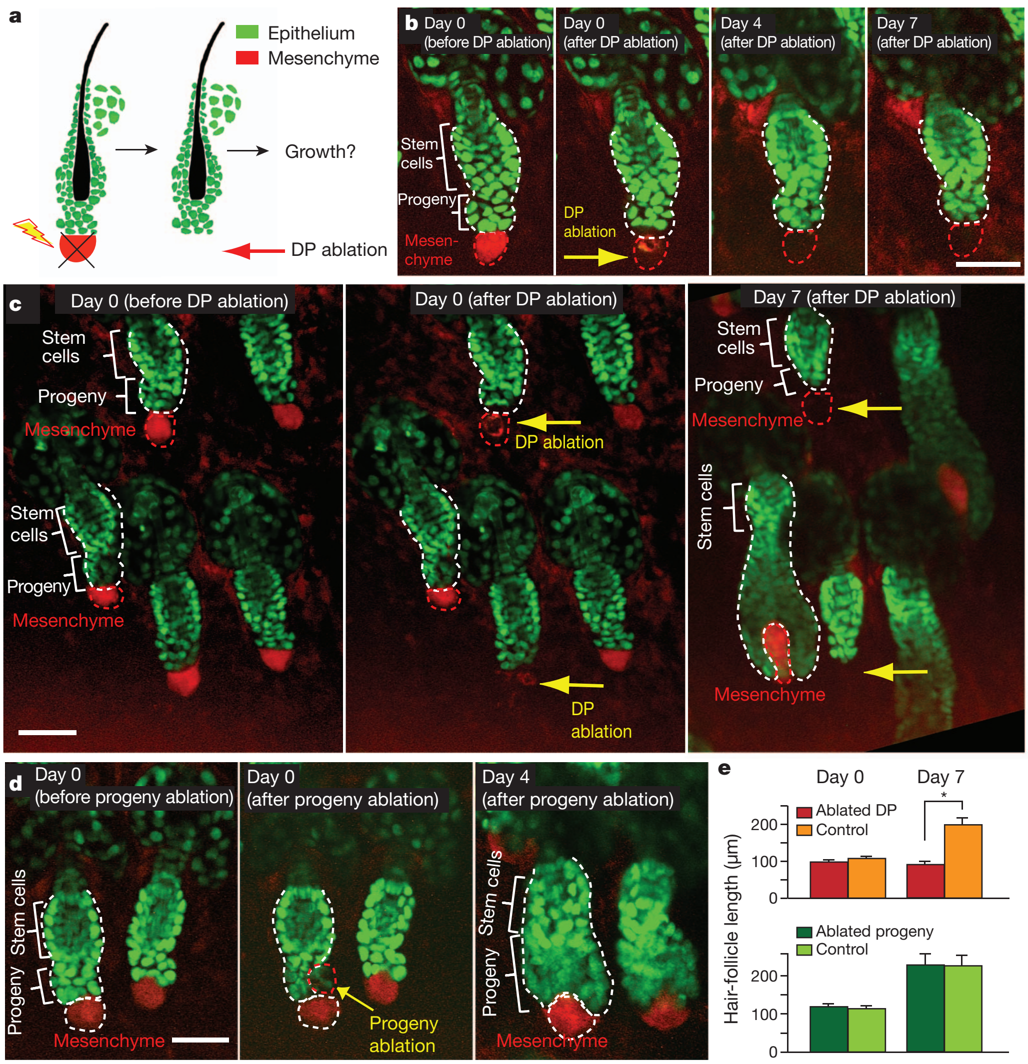


Figure 3 | Ablation of the mesenchymal dermal papilla impairs initiation of hair regeneration. **a**, Set-up for the experiment using laser-induced cell ablation of dermal papilla cells to test the requirement of the mesenchyme for hair growth. **b**, High-magnification optical sections of four time points of live hair follicles whose dermal papilla was ablated (approximately telogen phase). **c**, Low-magnification optical sections of three time points show a group of follicles in which only a few follicles had the dermal papilla ablated (yellow arrows). **d**, Low-magnification optical sections of three time points show two follicles of which one had the progeny partially ablated (yellow arrow). **e**, Quantification of the growth of hair follicles (measured as the total length of the hair follicle) with ablated dermal papilla (top) or partially ablated progeny compartments (bottom) compared to intact control follicles. Data are expressed as mean ± s.e.m. ($n = 8$–10; asterisk, $P < 0.0001$). Scale bars, 50 μm.

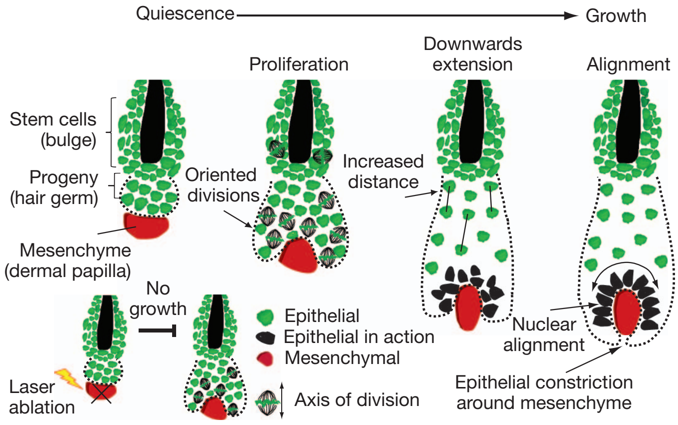


Figure 4 | The cellular mechanisms that participate in hair regeneration. During the initial stages of hair regeneration, the stem cell progeny is the first compartment that starts to proliferate. Cell divisions are also detected within the bulge, although the number of divisions is lower than in the stem cell progeny. Divisions in the progeny compartment are oriented along the axis of hair follicle growth, whereas orientations of divisions in the bulge are more random. The hair follicle undergoes a downwards extension in which the distance between nuclei increases within the progeny but not within the stem cell compartment. The epithelial nuclei surrounding the mesenchymal dermal papilla realign and constrict around the mensenchyme. Ablation of the mesenchyme results in impaired hair follicle growth.

retain the ability to host other tissue cell types such as melanocytes (Supplementary Fig. 9). To test whether failure to initiate hair growth was due to dermal papilla ablation rather than to collateral injury of the stem cell progeny, we targeted only cells within the progeny, leaving the dermal papilla intact. Hair follicles in which the progeny was injured recovered and grew at a similar rate to their surrounding uninjured follicles (Fig. 3d, e). These results indicate that the mesenchymal dermal papilla is required for hair growth initiation, and underscores the importance of epithelial–mesenchymal interactions for stem cell activation and regeneration.

In summary, we have imaged hair-follicle stem cells and their progeny in an uninjured living mammal, over a long period of time, with high spatial and temporal resolution. We have shown that hair regeneration involves a range of dynamic cellular behaviours and relies on the presence of the mesenchyme (Fig. 4). Given that components of the stem cell microenvironment are conserved across different tissues, our findings will probably be relevant to other tissues as well. More broadly, understanding how stem cell and progeny behaviour is regulated in physiological conditions may be necessary to advance our use of stem cells in regenerative medicine, and to uncover the cellular mechanisms that go awry in cancer and other diseases.

METHODS SUMMARY

K14H2BGFP, K14H2BGFP/Lef1RFP, K5tta and pTREH2BGFP transgenic mice were obtained from the Fuchs Laboratory and generated as previously described[3,21]. All studies and procedures involving animal subjects were approved by the Institutional Animal Care and Use Committee of the Yale School of Medicine and were conducted in accordance with the approved animal handling protocol. Mice between 3 and 4 weeks of age were anaesthetized by intraperitoneal injection, and the skin in the ear and head was depilated. An adjustable mounting spatula was inserted though the ear canal, and the skin connecting the ear to the head was flattened and dry mounted with a coverslip. Mice were maintained on a heating stage and provided with vaporized anaesthetic through a nose cone for the course of the live imaging session.

All mice were imaged with a LaVision TriM Scope II (LaVision Biotec) two-photon laser scanning microscope using a ×20 water immersion lens (N.A. 1.0; Olympus). Excitation (GFP, 940 nm; RFP, 1040 nm) and ablation (900 nm) were performed using a Chameleon Vision II (Coherent) laser. Three-dimensional stacks consisting of 50 optical sections in 2-μm steps along the *z* axis, providing a depth of approximately 100 μm below the surface of the skin, were acquired every 5 min. Image processing, measurements, assembly and editing of time-lapse movies were performed using Image J (NIH Image) or Volocity (Perkin Elmer). An unpaired Student's *t*-test was applied in the two-group statistical analysis and *P* values of less than 0.05 were considered significant. Histological and immunofluorescence analysis of frozen skin sections was performed following standard protocols as described previously[6].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 December 2011; accepted 8 May 2012.
Published online 1 July 2012.

1. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132,** 598–611 (2008).
2. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61,** 1329–1337 (1990).
3. Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303,** 359–363 (2004).
4. Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. & Barrandon, Y. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104,** 233–245 (2001).
5. Jahoda, C. A., Horne, K. A. & Oliver, R. F. Induction of hair growth by implantation of cultured dermal papilla cells. *Nature* **311,** 560–562 (1984).
6. Greco, V. *et al.* A two-step mechanism for stem cell activation during hair regeneration. *Cell Stem Cell* **4,** 155–169 (2009).
7. Xie, Y. *et al.* Detection of functional haematopoietic stem cell niche using real-time imaging. *Nature* **457,** 97–101 (2009).
8. Lo Celso, C. *et al.* Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457,** 92–96 (2009).
9. Uchugonova, A., Hoffman, R. M., Weinigel, M. & Koenig, K. Watching stem cells in the skin of living mice noninvasively. *Cell Cycle* **10,** 2017–2020 (2011).
10. Ra, H. *et al. In vivo* imaging of human and mouse skin with a handheld dual-axis confocal fluorescence microscope. *J. Invest. Dermatol.* **131,** 1061–1066 (2011).
11. Watt, F. M. & Jensen, K. B. Epidermal stem cell diversity and quiescence. *EMBO Mol. Med.* **1,** 260–267 (2009).
12. Fuchs, E. The tortoise and the hair: slow-cycling cells in the stem cell race. *Cell* **137,** 811–819 (2009).
13. Greco, V. & Guo, S. Compartmentalized organization: a common and required feature of stem cell niches? *Development* **137,** 1586–1594 (2010).
14. Waghmare, S. K. *et al.* Quantitative proliferation dynamics and random chromosome segregation of hair follicle stem cells. *EMBO J.* **27,** 1309–1320 (2008).
15. Müller-Röver, S. *et al.* A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *J. Invest. Dermatol.* **117,** 3–15 (2001).
16. Wallingford, J. B., Fraser, S. E. & Harland, R. M. Convergent extension: the molecular control of polarized cell movement during embryonic development. *Dev. Cell* **2,** 695–706 (2002).
17. Zhang, Y. V., Cheong, J., Ciapurin, N., McDermitt, D. J. & Tumbar, T. Distinct self-renewal and differentiation phases in the niche of infrequently dividing hair follicle stem cells. *Cell Stem Cell* **5,** 267–278 (2009).
18. Petersson, M. *et al.* TCF/Lef1 activity controls establishment of diverse stem and progenitor cell compartments in mouse epidermis. *EMBO J.* **30,** 3004–3018 (2011).
19. Rubin, L. & Saunders, J. W. Jr. Ectodermal–mesodermal interactions in the growth of limb buds in the chick embryo: constancy and temporal limits of the ectodermal induction. *Dev. Biol.* **28,** 94–112 (1972).
20. Plikus, M. V. *et al.* Cyclic dermal BMP signalling regulates stem cell activation during hair regeneration. *Nature* **451,** 340–344 (2008).
21. Rendl, M., Lewis, L. & Fuchs, E. Molecular dissection of mesenchymal-epithelial interactions in the hair follicle. *PLoS Biol.* **3,** e331 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Horwich, V. Reinke and A. Giraldez for critical discussion of the work. We thank S. King for feedback on the manuscript, and E. Fuchs for the K14H2BGFP, Lef1RFP and pTREH2BGFP mice and A. Glick for the K5tta mice. P.R. is supported by the James Hudson Brown – Alexander Brown Coxe postdoctoral fellowship. G.Z. is a New York Stem Cell Foundation–Druckenmiller Fellow. E.D. is supported by the Cell and Molecular Biology Training Grant of the National Institute of General Medical Sciences, US National Institutes of Health, Public Health Service. This work was supported in part by the American Skin Association, the American Cancer Society and the Yale Rheumatologic Disease Research Core Center grant NIH P30AR053495.

Author Contributions P.R. and V.G. designed experiments and wrote the manuscript, and P.R. performed the experiments and analysed the data. E.D. performed two-photon laser-scanning timelapses. G.Z. performed immunostainings. I.S. set up the mouse colonies and staining protocols. D.G. and A.H. assisted on initial intravital imaging set-up. D.G. performed data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.G. (valentina.greco@yale.edu).

METHODS

Transgenic mice. K14H2BGFP, K14H2BGFP/Lef1RFP, K5tta and pTREH2BGFP transgenic mice were generated as previously described[3,21]. All studies and procedures involving animal subjects were approved by the Institutional Animal Care and Use Committee of the Yale School of Medicine and were conducted in accordance with the approved animal handling protocol.

***In vivo* imaging.** Three-week-old mice were anaesthetized with intraperitoneal injection of ketamine and xylazine, and the skin around the head region was shaved using a mechanical trimmer and depilatory cream. The mouse was placed on a heated stage, and the head and the ear were supported by a custom-made stage. A glass coverslip was placed against the skin in the junction region between the head and the ear. Image stacks of the skin were acquired with a LaVision TriM Scope II (LaVision Biotec) microscope equipped with a Chameleon Vision II (Coherent) two-photon laser. A laser beam (at 940 nm for GFP and 1040 nm for RFP, respectively) was focused through a $\times$20 water immersion lens (N.A. 1.0; Olympus) and scanned with a field of view of 0.25 to 0.5 mm^2 at 600 Hz. Serial optical sections were acquired in 2–3-μm steps to image a total depth of ~100 μm of tissue in 5-min intervals. Several phases covering the transition from quiescent to growth stages were analysed (telogen to anagen phases). Distinctive inherent landmarks in the skin were used to navigate back to the original field of view and visualize the same follicles in separate experiments. Anaesthesia was maintained throughout the course of the experiment with vaporized isofluorane delivered by a nose cone.

Three-dimensional two-photon laser ablation. Laser ablation was carried out with the same optics as for acquisition, essentially as described previously[22]. A 900-nm laser beam was used to scan a 10 μm^2 area and ablation was achieved using 25% laser power for 1 s. Ablation parameters were adjusted according to the depth of the target (30–80 μm).

Image analysis. Raw image stacks (512 $\times$ 512 to 1024 $\times$ 1024 pixels in the x–y plane and 2–3-μm voxel depth; typically 30–50 optical sections) were imported into ImageJ (NIH Image) for further analysis. Optical planes from sequential time points were manually realigned to compensate for minor tissue drift during the time course of the acquisition. The contrast was adjusted accordingly and selected optical planes or z-projections of sequential optical sections were used to assemble time-lapse movies. Three-dimensional rendering and cell tracking was performed using the Volocity software package (Perkin Elmer).

Statistical analysis. Data are expressed as mean $\pm$ s.e.m. An unpaired Student's t-test was used to analyse data sets with two groups. For all analyses, a P value of less than 0.05 was accepted as indicating a significant difference. Statistical calculations were performed using the Prism software package (GraphPad).

Histology and immunolabelling. Sections frozen in Tissue-Tek OCT Compound (Sakura Finetek) were fixed with 4% formaldehyde for 10 min at room temperature (22 °C) and treated as described previously[6]. Antibodies were used at the following dilutions: Ncam (Chemicon, rat 1:100), CD34 (eBioscience, rat 1:50), Pcadherin (R&D, goat 1:100) and Trp2 (Santa Cruz, goat 1:100).

22. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472,** 51–56 (2011).

doi:10.1038/nature11217

RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis

Min Yu[1,2]*, David T. Ting[1]*, Shannon L. Stott[1], Ben S. Wittner[1], Fatih Ozsolak[3], Suchismita Paul[1], Jordan C. Ciciliano[1], Malgorzata E. Smas[1], Daniel Winokur[1], Anna J. Gilman[1], Matthew J. Ulman[1], Kristina Xega[1], Gianmarco Contino[1], Brinda Alagesan[1], Brian W. Brannigan[1], Patrice M. Milos[3], David P. Ryan[1], Lecia V. Sequist[1], Nabeel Bardeesy[1], Sridhar Ramaswamy[1], Mehmet Toner[1], Shyamala Maheswaran[1] & Daniel A. Haber[1,2]

Circulating tumour cells (CTCs) shed into blood from primary cancers include putative precursors that initiate distal metastases[1]. Although these cells are extraordinarily rare, they may identify cellular pathways contributing to the blood-borne dissemination of cancer. Here, we adapted a microfluidic device[2] for efficient capture of CTCs from an endogenous mouse pancreatic cancer model[3] and subjected CTCs to single-molecule RNA sequencing[4], identifying *Wnt2* as a candidate gene enriched in CTCs. Expression of WNT2 in pancreatic cancer cells suppresses anoikis, enhances anchorage-independent sphere formation, and increases metastatic propensity *in vivo*. This effect is correlated with fibronectin upregulation and suppressed by inhibition of MAP3K7 (also known as TAK1) kinase. In humans, formation of non-adherent tumour spheres by pancreatic cancer cells is associated with upregulation of multiple WNT genes, and pancreatic CTCs revealed enrichment for WNT signalling in 5 out of 11 cases. Thus, molecular analysis of CTCs may identify candidate therapeutic targets to prevent the distal spread of cancer.

We established the ^{Hb}CTC-Chip microfluidic capture platform[1] to isolate CTCs from blood samples of a genetically engineered mouse pancreatic cancer model[2], comparing their expression profile to that of simultaneously collected primary tumour and metastatic ascites cells (Fig. 1a and Supplementary Fig. 1). Captured cells were stained for epithelial cytokeratin (CK) and the leukocyte marker CD45, with CTCs scored as $CK^+/CD45^-$ (Fig. 1b). Eight control tumour-free mice had a median of 1.3 $CK^+/CD45^-$ cells per 100 µl of blood (range, 0.2–5.6; mean ± s.d., 1.8 ± 1.7) under the selected imaging parameters. Applying a threshold of ≥ 6 cells per 100 µl, 7 out of 8 tumour-bearing mice were positive for CTCs (median, 31 cells per 100 µl; range, 2–547; mean ± s.d., 115 ± 188) (Fig. 1b, c). Most CTCs were captured as single cells, although clusters of 5 to 10 cells were seen in ~50% of mouse blood samples. The on-chip purity of captured CTCs ranged from 0.1–6%, due to non-specific binding (NSB) by leukocytes (Supplementary Table 1).

Given the minute amounts and variable purity of CTCs, we applied a sub-microgram RNA-based sequencing method using a single-molecule 'next generation' platform[3], to derive a digital gene expression (DGE) profile. We processed each blood sample through paired EPCAM- and mock IgG-functionalized ^{Hb}CTC-Chips, allowing digital subtraction of matched NSB leukocyte reads from each CTC-enriched DGE profile. Using the DEGseq statistical model[5] applied to the MA-plot with a false discovery rate (FDR) of 0.10, we identified 361 transcripts in mouse MPANC-9 cells that were increased more than twofold in anti-EPCAM- versus IgG-chips, and absent in blood from non-tumour-bearing mice (Fig. 1d). This CTC candidate gene set was then compared with DGE profiles of 12 mouse and 15 human primary pancreatic tumours, versus normal mouse and human pancreata[6], yielding a set of 9 tumour-associated genes whose expression was markedly increased in CTCs (Supplementary Table 2). One of these, *Wnt2*, was found to be consistently enriched in pancreatic cancers in the Oncomine database, leading us to pursue this gene as a prototype CTC-enriched transcript. RNA sequencing of metastatic cells in the mouse ascites demonstrated enrichment for *Wnt2*. A second mouse, MPANC-10, also demonstrated *Wnt2* expression in both CTCs and metastatic ascites cells. The lower number of *Wnt2* messenger RNA reads in MPANC-10 cells was correlated with the fewer cytokeratin transcript reads, indicating that both transcripts track with the number of CTCs in the blood specimen (Supplementary Table 3).

Given the absence of antibodies for cellular imaging of WNT2, we developed fluorescent RNA *in-situ* hybridization (RNA-ISH) to verify *Wnt2* expression in CTCs (Supplementary Figs 2 and 3). RNA-ISH analysis of CTCs using dual staining for *Wnt2* and cytokeratin mRNAs identified *Wnt2* transcripts in 64% of cytokeratin-expressing cells in 3 out of 4 mice. A comparable *Wnt2* mRNA signal was observed in metastatic cells from ascites of mice bearing pancreatic tumours (Fig. 1e). In contrast, within primary tumour specimens, *Wnt2* mRNA expression was only detectable in very small localized clusters of cells (1–5% of all cells) in 8 out of 14 primary tumour specimens (Fig. 1f). Histological analysis of rare *Wnt2*-positive cells within pancreatic ductal adenocarcinomas (PDACs) did not reveal any obvious distinction from other tumour cells. The small number of *Wnt2*-positive cells in primary tumour specimens was consistent with DGE analysis, which showed rare *Wnt2* RNA reads in both mouse (8/12) and human (9/15) PDACs. Thus, *Wnt2*-positive cells are present within both CTC and metastatic ascites, and represent a rare subset of the primary tumour population.

To test the functional consequences of WNT2 expression *in vivo*, lentiviral constructs were introduced into NB508 PDAC cells, which lack detectable endogenous WNT2 expression (Supplementary Fig. 4). Subcutaneous engraftment of green fluorescent protein (GFP)-tagged Wnt2-NB508 cells produced significantly more lung metastases in nude mice, compared with vector-transduced cells, despite the larger size of vector-expressing primary tumours ($n = 8$, $P < 0.05$) (Fig. 2a and Supplementary Fig. 4). EPCAM-captured CTCs showed a modest increase in numbers in mice bearing WNT2-expressing tumours, but this did not reach statistical significance ($P = 0.25$, Fig. 2b), even after normalization for the smaller size of WNT2-expressing tumours ($P = 0.17$, Supplementary Fig. 4). Thus, WNT2 expression may increase the metastatic potential of circulating cancer cells, without markedly increasing their generation from primary tumours. Consistent with this hypothesis, direct intravascular inoculation of cells into the tail vein of syngeneic mice, which bypasses the vascular invasion step, showed a marked increase in lung metastases for Wnt2-NB508 cells (5/6 Wnt2-transduced versus 0/6 vector, $P < 0.05$) (Fig. 2c and Supplementary Fig. 4).

[1]Massachusetts General Hospital Cancer Center, Center for Engineering in Medicine, and Departments of Medicine and Surgery, Harvard Medical School, Boston, Massachusetts 02114, USA. [2]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. [3]Helicos BioSciences Corporation, One Kendall Square, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

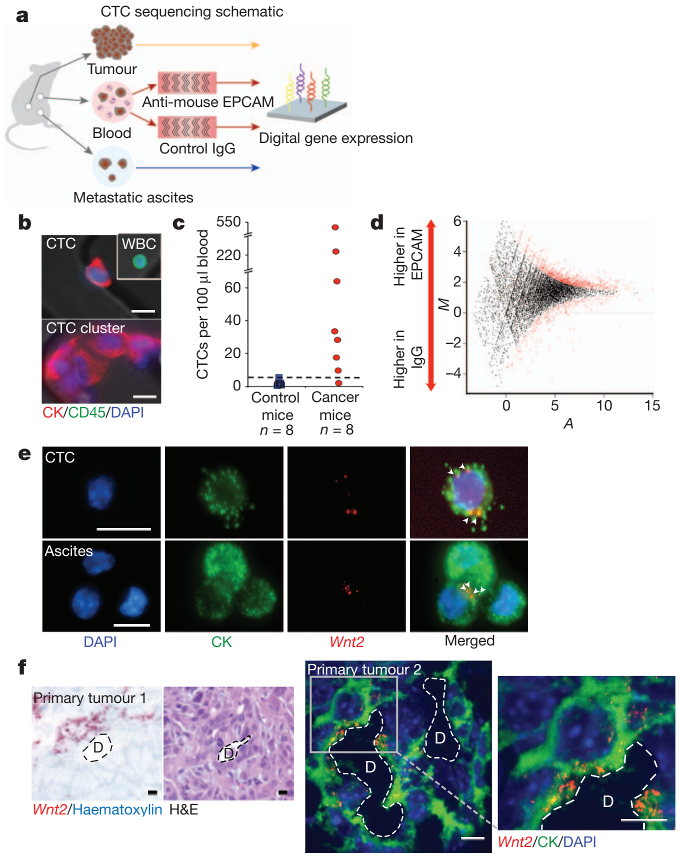


Figure 1 | Analysis of mouse pancreatic CTCs identifies *Wnt2* as a candidate CTC gene. **a**, Schematic representation of strategy for DGE of RNA isolated from primary mouse pancreatic tumour, metastatic ascites, and CTC-enriched blood. **b**, Immunofluorescence staining of mouse pancreatic CTC, leukocyte (WBC), and CTC cluster captured on ^{Hb}CTC-Chip (DAPI, blue; cytokeratin (CK), red; CD45, green). **c**, Quantification of CK^+ cells captured from control mice and pancreatic cancer bearing mice (dashed line, threshold of ≥ 6 CK^+ cells per 100 μl). **d**, DEGseq *M*–*A* plot of RNA sequence reads from mouse (MPANC-9) CTCs, comparing anti-EPCAM-captured versus IgG-coated ^{Hb}CTC-Chips. (*M*, $\log_2$ fold change, *A*, averaged $\log_2$ reads, red dots, differentially expressed genes) **e**, RNA-ISH of CTCs and metastatic ascites cells co-expressing CK8+18 (CK, green) and *Wnt2* (red) transcripts. White arrowheads highlight *Wnt2* signals. **f**, RNA-ISH of primary pancreatic tumours showing a small cluster of *Wnt2* mRNA expressing cells near a pancreatic duct (D) (*Wnt2*, red dots; haematoxylin, light blue). Haematoxylin and eosin (H&E) stained serial section is shown (right). A second primary tumour imaged under fluorescence demonstrating *Wnt2* expression in a subpopulation of tumour cells; high magnification shown (right, *Wnt2*, red; CK8+18, green; DAPI, blue). Scale bars, 10 μm.

Modelling the effects of ectopic WNT2 expression on NB508 PDAC cells grown *in vitro* showed marked enhancement of anchorage-independent growth, without affecting cellular proliferation, migration or invasion assayed under standard conditions (Supplementary Figs 5, 6). Tumour sphere formation under non-adherent conditions in serum-free media supplemented with growth factors[7,8] was markedly increased by WNT2 expression, with respect to sphere numbers ($P < 0.05$) and size ($P < 0.005$) (Fig. 2d and Supplementary Fig. 7). Anoikis, the induction of apoptosis in epithelial cells resulting from loss of basement membrane attachment, was attenuated following WNT2 expression in NB508 cells (Fig. 2e). Resistance to anoikis is thought to be essential for epithelial cell survival in the bloodstream, and epithelial to mesenchymal transition (EMT) has been postulated to mediate this effect. We therefore tested for induction of characteristic mesenchymal markers by WNT2 at both RNA and protein levels, by DGE analysis and western blotting, respectively. Among classical EMT markers, we only detected induction of fibronectin (FN1), an extracellular protein implicated in cell–matrix interactions and cellular survival signals (Fig. 2e and Supplementary Figs 6, 8). Short hairpin RNA-mediated knockdown of *Fn1* suppressed the ability of ectopic WNT2 to enhance tumour sphere formation ($n = 6$, $P < 0.05$, Fig. 2f), indicating that *Fn1* contributes to WNT2-mediated anoikis resistance. Two transcription factors, MYCN and the known FN1 regulator ETV4, were induced by WNT2, with their knockdown suppressing FN1 expression, indicating that they contribute to mediating the WNT2 effect (Supplementary Fig. 9).

WNT2 may signal through either canonical (β-catenin/TCF dependent) or non-canonical pathways. Whereas canonical WNT signalling expression signatures have been developed, there are no such signatures for non-canonical WNT pathways. We therefore generated a customized non-canonical WNT signalling gene set (Supplementary Table 4)[9] and used this to interrogate the CTC DGE data. Significant enrichment was observed for components of the non-canonical WNT signalling pathway in mouse pancreatic CTCs, metastatic ascites cells, and Wnt2-NB508 cells grown under non-adherent sphere-forming conditions (Supplementary Figs 10, 11 and Supplementary Table 5). Similar enrichment was noted for a WNT- and TGF-β-driven signature[10] (Supplementary Table 6).

No such enrichment was evident for signatures of canonical WNT signalling in CTCs and metastatic ascites cells, although Wnt2-NB508 cells showed a mild increase in canonical WNT signalling (Supplementary Table 6, Supplementary Fig. 12).

We tested a panel of inhibitors of WNT-related pathways to identify small-molecule inhibitors capable of suppressing the WNT2 effect on anoikis. Among these, 5*Z*-7-oxozeaenol, an inhibitor of MAP3K7, also known as TAK1 (TGF-β activated kinase 1)[11], was remarkable in completely abrogating WNT2-induced tumour spheres, without suppressing baseline sphere formation (Fig. 3a and Supplementary Figs 12, 13). 5*Z*-7-oxozeaenol also suppressed FN1 expression in Wnt2-NB508 cells (Fig. 3b). To validate the effect of the TAK1 inhibitor in reversing the pro-survival phenotype conferred by WNT2, shRNA constructs targeting TAK1 were introduced into Wnt2-NB508 cells. TAK1 shRNAs suppressed FN1 expression in Wnt2-NB508 cells with an associated increase in cleaved caspase-3 (Fig. 3c and Supplementary Fig. 9d). Effective suppression of WNT2-induced tumour sphere formation was observed and correlated with the degree of TAK1 knockdown by the shRNA constructs (Fig. 3d). Consistent with these *in vitro* findings, the generation of metastases following tail vein inoculation of WNT2-expressing cells was suppressed as a function of shRNA knockdown of TAK1 (Fig. 3e and Supplementary Fig. 14). Together, these results support the role of WNT2 in mediating metastasis-associated survival signals, which depend, at least in part, on a TAK1–FN1 signalling axis.

To extend our analyses to human pancreatic cancer, we captured CTCs from blood specimens of patients with metastatic pancreatic cancer using the ^{Hb}CTC-Chip and stained the cells with a combination of antibodies against cytokeratin and intracellular EPCAM residues (Fig. 4a). Under conditions optimized for human pancreatic CTC analysis, 10 healthy controls had a median 1.2 $CK^+EPCAM^+/CD45^-$ cells (range, 0 to 2.2; mean ± s.d., 1.3 ± 0.8). Of 15 patients with metastatic pancreatic cancer, 11 had CTC counts above threshold (≥ 3 $CK^+EPCAM^+/CK45^-$ cells per ml), with a median 11 cells per ml (range, 1 to 57; mean ± s.d., 18 ± 18; on-chip median purity, 0.5%) (Fig. 4b and Supplementary Table 7). We tested for *WNT2* expression in human pancreatic CTCs using RNA-ISH identifying *WNT2* transcripts in 23 out of 66 (35%) CK-positive CTCs from 2 out of 8 metastatic PDAC patients (Fig. 4c). To derive DGE profiles for human pancreatic CTCs, we followed the strategy defined for mouse CTC analyses, analysing blood specimens from 12 patients and 4 healthy controls using parallel anti-EPCAM versus mock IgG-functionalized ^{Hb}CTC-Chips, followed by low-quantity RNA-based single-molecule sequencing analysis. No individual WNT transcript was consistently enriched across all CTC-enriched populations (FDR 0.10), but among

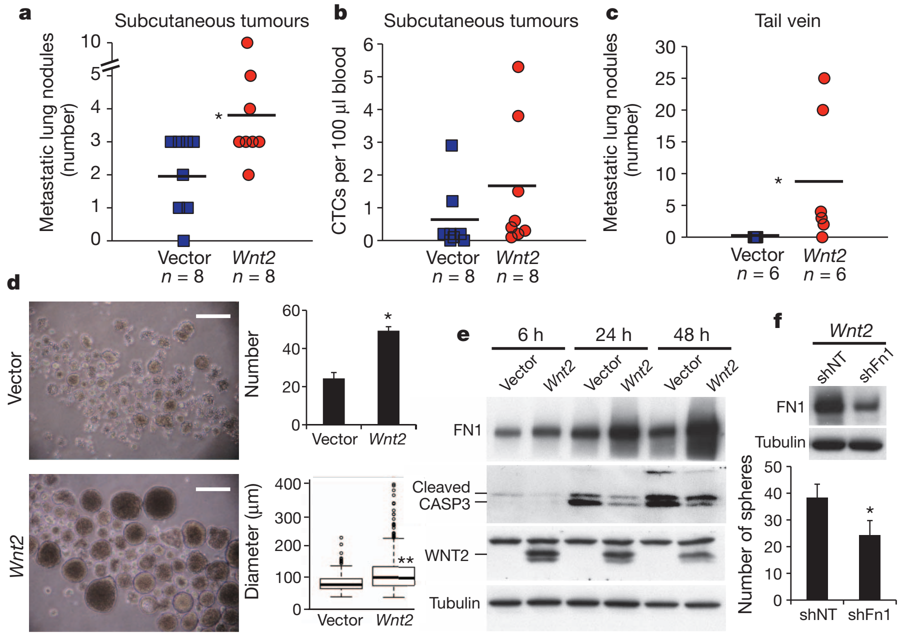


Figure 2 | WNT2 promotes anchorage-independent cell survival and pancreatic cancer cell metastasis. **a**, Number of GFP-positive metastatic nodules in the lungs of mice bearing subcutaneous tumours established with vector- or WNT2-expressing NB508 cells. **b**, Number of GFP-positive CTCs captured from the blood of mice described above. **c**, Number of GFP-positive lung metastatic nodules in mice, following tail vein injection with vector- or Wnt2-NB508 cells tagged with both GFP and luciferase. **d**, Representative images of tumour spheres formed by vector- or Wnt2-NB508 cells plated at 1,000 cells per well (scale bar, 250 μm). Quantification of both tumour sphere number and diameter is shown plated at 100 cells per well ($n = 3$). **e**, Immunoblotting analysis of vector- and Wnt2-NB508 cells at time intervals following plating under non-adherent conditions. **f**, Suppression of tumour spheres formed by Wnt2-NB508 cells following infection with lentivirus encoding shRNA targeting *Fn1* compared with non-target (NT) shRNA ($n = 3$). Effectiveness of knockdown is shown by western blot above. (mean ± s.d.; $^{*}P < 0.01$, $^{**}P < 0.001$).

11 patients with CTC counts above threshold, 5 (45%) had significant enrichment of non-canonical Wnt signalling pathway components and 8 (73%) showed enrichment for the WNT- and TGF-β-driven profile (Supplementary Table 8). No such enrichment was present in 4 healthy controls or in one patient without detectable CTCs (Supplementary Tables 9–11).

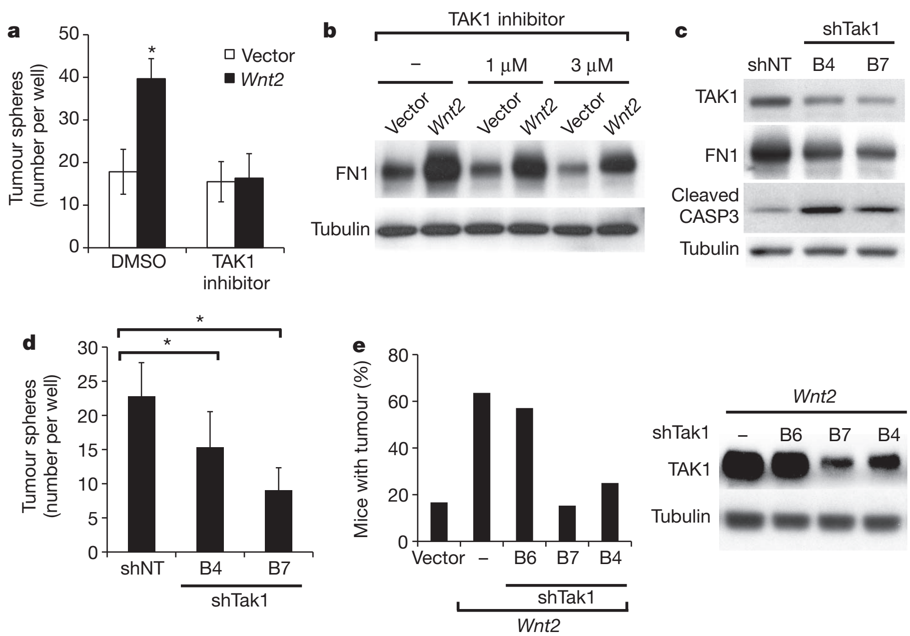


Figure 3 | Association of WNT2-prosurvival phenotype with non-canonical WNT signalling and inhibition by suppression of MAP3K7 (TAK1). **a**, Suppression of Wnt2-induced, but not baseline, tumour sphere formation by NB508 cells, following treatment with 5*Z*-7-Oxozeaenol, a small molecule inhibitor of TAK1 ($n = 6$). **b**, Immunoblot showing suppression of FN1 expression following treatment of non-adherent cultures of Vec- and Wnt2-NB508 cells with increasing concentrations of 5*Z*-7-Oxozeaenol. **c**, Immunoblot analysis of TAK1, FN1 and cleaved caspase 3 expression in Wnt2-NB508 cells following lentiviral infection with two different shRNAs targeting *Tak1* or non-target (NT). **d**, Enumeration of tumour spheres formed by cells in c ($n = 6$). **e**, Percentage of mice with detectable metastases by bioluminescence imaging for luciferase-producing tumour cells in mice 6 weeks following tail vein injection of Vec-NB508, Wnt2-NB508, and Wnt2-NB508 cells infected with three different shRNAs targeting *Tak1* ($n = 12$, 11, 14, 13 and 12, respectively). Immunoblot of TAK1 knockdown by shRNAs B6, B7, and B4 is shown (right). (mean ± s.d., $^{*}P < 0.05$)

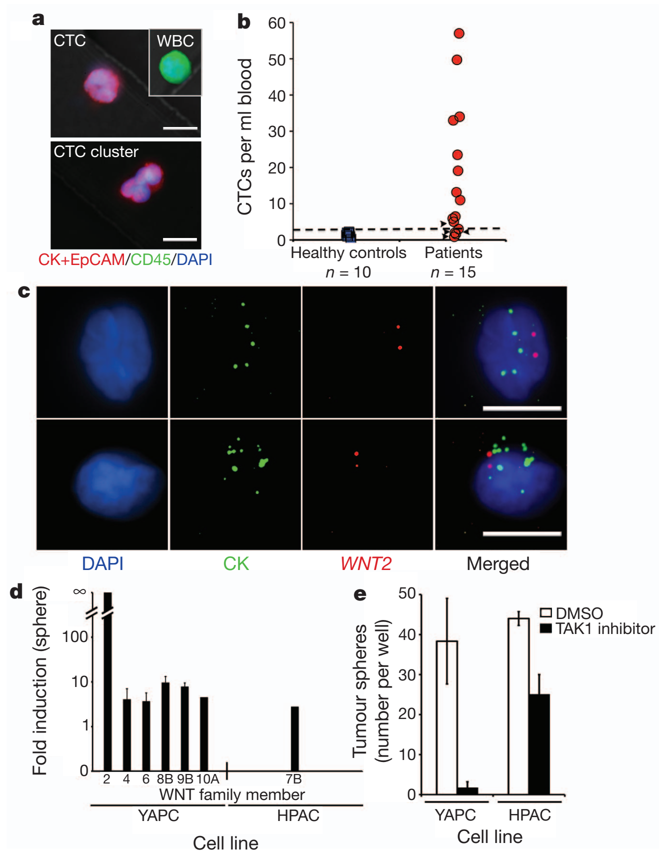


Figure 4 | Detection of WNT2 mRNA expression and non-canonical WNT signature in human pancreatic CTCs. **a**, Immunofluorescence staining of human pancreatic CTC, leukocyte (WBC), and CTC cluster captured on anti-human EPCAM ^{Hb}CTC-Chip (DAPI nuclear stain, blue; CK and EPCAM cocktail, red; CD45, green). **b**, Enumeration of CK+EPCAM positive human CTCs (CTCs/ml) captured from patients with metastatic pancreatic cancer. Arrowheads indicate lower CTC numbers in patients responding to therapy. Blood samples from healthy donors were used to establish the threshold of ≥ 3 $CK^{+}EPCAM^{+}/CK45^{-}$ cells/ml (dashed line). **c**, RNA-ISH analysis of human pancreatic CTCs, showing co-expression of mRNAs for cytokeratins (CK) 7, 8, 18, 19 and 23 (pooled probes in green) and *WNT2* (red). (DAPI, blue). Scale bars, 10 μm. **d**, Induction of WNT transcripts in pancreatic cancer cells grown as non-adherent tumour spheres compared to standard conditions (fold increase). **e**, Quantification of tumour spheres with or without TAK1 inhibitor (3μM 5Z-7-Oxozeanol). ($n = 3$; mean ± s.d.)

To test whether multiple WNT genes contribute to anoikis resistance in human pancreatic cancer, we measured expression of all known family members in six cell lines grown under non-adherent tumour sphere conditions. Five out of six cell lines showed upregulation of multiple WNT genes following shift from two-dimensional to three-dimensional culture, an effect that was correlated with sensitivity to the TAK1 inhibitor and to *TAK1* shRNA (Fig. 4d, e and Supplementary Figs 15–17, Supplementary Table 12). Thus, in human pancreatic cancer cells, the induction of multiple WNT family members seems to circumvent anoikis following loss of adherence, an effect suppressed by inhibition of TAK1.

In summary, we have applied microfluidic CTC isolation from a genetically engineered mouse model to generate an unbiased RNA sequencing profile for CTC-specific expression patterns. The cellular pathways activated during the haematogenous spread of human pancreatic cancer are likely to be complex, and our analysis is limited by technological hurdles inherent in the molecular characterization of partially purified CTC populations. Nonetheless, our results suggest that non-canonical WNT signalling pathways may contribute to metastasis in human pancreatic cancer. The effectiveness of TAK1 inhibition in suppressing this effect identifies a novel potential drug target for metastasis suppression.

METHODS SUMMARY

Mice with pancreatic cancer express Cre driven by *Pdx1* or *P48*, along with $LSL\text{-}Kras^{G12D}$, and $Tp53^{lox/+}$ or $Tp53^{lox/lox}$ as previously described[3]. The mouse pancreatic tumour cell lines were generated from the resultant cancers. Mouse blood was obtained via cardiocentesis and was diluted 1:1 with PBS with 10 mM EDTA pH 7.4. Human blood for pancreatic CTC analysis was obtained after informed consent as per IRB protocol (05-300) at the Massachusetts General Hospital (MGH). ^{Hb}CTC-Chip production was as previously described[1]. Mouse ^{Hb}CTC-Chips were functionalized with biotinylated anti-mouse EPCAM (CD326) antibody (Biolegend 118204) or isotype control biotinylated Rat IgG2a (Biolegend 400504). Human ^{Hb}CTC-Chips were functionalized with biotinylated polyclonal goat anti-human EPCAM (R&D Systems BAF960) or control goat IgG antibody (R&D Systems BAF108). RNA was isolated from the devices using a modified protocol from the Qiagen RNeasy MinElute kit (Qiagen). Helicos single-molecule sequencing was performed as previously described[4]. QuantiGene ViewRNA *in-situ* hybridization on ^{Hb}CTC-Chip was performed with a modified protocol from the manufacturer (Affymetrix).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 July 2011; accepted 8 May 2012.
Published online 1 July 2012.

1. Yu, M., Stott, S., Toner, M., Maheswaran, S. & Haber, D. A. Circulating tumor cells: approaches to isolation and characterization. *J. Cell Biol.* **192,** 373–382 (2011).
2. Stott, S. L. *et al.* Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. *Proc. Natl Acad. Sci. USA* **107,** 18392–18397 (2010).
3. Bardeesy, N. *et al.* Both $p16^{Ink4a}$ and the $p19^{Arf}$-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl Acad. Sci. USA* **103,** 5947–5952 (2006).
4. Ozsolak, F. *et al.* Amplification-free digital gene expression profiling from minute cell quantities. *Nature Methods* **7,** 619–621 (2010).
5. Wang, L., Feng, Z., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26,** 136–138 (2010).
6. Ting, D. T. *et al.* Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* **331,** 593–596 (2011).
7. Reynolds, B. A. & Weiss, S. Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell. *Dev. Biol.* **175,** 1–13 (1996).
8. Dontu, G. *et al. In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17,** 1253–1270 (2003).
9. Katoh, M. WNT signaling pathway and stem cell signaling network. *Clin. Cancer Res.* **13,** 4042–4045 (2007).
10. Hoshida, Y. *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* **69,** 7385–7392 (2009).
11. Ninomiya-Tsuji, J. *et al.* A resorcylic acid lactone, 5*Z*–7-oxozeaenol, prevents inflammation by inhibiting the catalytic activity of TAK1 MAPK kinase kinase. *J. Biol. Chem.* **278,** 18485–18490 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Koris and the MGH clinical research coordinators for help with clinical studies, J. Gentry for computer informatics support, L. Libby for mouse studies, J. Walsh for microscopy expertise, Z. Nakamura for technical support, D. Jones for sequencing, and M. Rivera and S. Akhavanfard for RNA-ISH. This work was supported by Stand Up To Cancer (D.A.H., M.T., S.M.), Howard Hughes Medical Institute (D.A.H.), NIBIB Quantum Grant 5R01EB008047 (M.T.), NIH CA129933 (D.A.H.), Pancreatic Cancer Action Network – AACR Fellowship (D.T.T.), the Warshaw Institute for Pancreatic Cancer Research (D.T.T.), and the K12 Paul Calabresi Award for Clinical Oncology Clinical Research Career Development Program NIH 5K12CA87723-09 (D.T.T.).

Author Contributions M.Y., D.T.T., S.M. and D.A.H. designed and conducted the study, analysed data, and prepared the manuscript. S.L.S. and M.T. contributed to the microfluidic system. D.T.T., B.S.W., F.O., B.W.B. and P.M.M. performed Helicos digital gene expression analysis. B.S.W. and S.R. performed bioinformatic and statistical analysis. S.P., J.C.C., M.E.S., D.W., A.J.G., M.J.U. and K.X. provided technical assistance for microfluidic experiments. M.Y., D.T.T., B.W.B. and J.C.C. performed RNA-ISH analysis. D.T.T., D.P.R. and L.V.S. acquired clinical samples. G.C., B.A. and N.B. provided the mouse model and cell line. S.L.S., B.S.W., P.M.M., S.R. and M.T. commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.M. (maheswaran@helix.mgh.harvard.edu) or D.A.H. (haber@helix.mgh.harvard.edu).

METHODS

Mice, tumour tissues and cell lines. Mice with pancreatic cancer used in these experiments express Cre driven by *Pdx1* or *P48*, *LSL-Kras*^{G12D}, and *Tp53*$^{lox/+}$ or *Tp53*$^{lox/lox}$ as previously described[2]. Normal B6/129, FVB, and Nu/Nu mice were purchased from Jackson Laboratory. For cardiocentesis, animals were sedated with isofluorane, the chest wall was sterilized with ethanol and a skin incision was made above the rib cage. To avoid potential contamination from cancer cells in the peritoneum, a 23-gauge needle was inserted through the intercostal muscle of the left chest into the heart and approximately 1 ml of blood was draw into a syringe primed with 100 μl of PBS with 10 mM EDTA pH 7.4 (Gibco). Animals were then euthanized per animal protocol guidelines. Pancreatic tumours and normal tissues were extracted and separated using flash-freezing with liquid nitrogen for RNA extraction and fixation with 4% paraformaldehyde (PFA) followed by paraffin embedding (FFPE) for RNA-ISH and histochemistry. Ascites if present were sterilely removed with a 23G needle syringe and cells were pelleted by centrifugation. Pelleted cells were flash-frozen in liquid nitrogen for RNA extraction. Cell lines generated from mouse tumours and human pancreatic cancer cell lines were grown in RPMI-1640 + 10% FBS + 1% penicillin/streptomycin (Gibco/Invitrogen).

^{Hb}CTC-Chip production, preparation and blood processing. ^{Hb}CTC-Chips were manufactured on site at the MGH Cancer Center/BioMEMS Resource Facility, and chips were functionalized using biotinylated antibodies as previously described[1]. Mouse ^{Hb}CTC-Chips were functionalized with biotinylated anti-mouse EPCAM (CD326) antibody (Biolegend 118204) or isotype control biotinylated Rat IgG2a (Biolegend 400504). Human ^{Hb}CTC-Chips were functionalized with biotinylated polyclonal goat anti-human EPCAM (R&D Systems BAF960) or control goat IgG antibody (R&D Systems BAF108).

For mouse samples, approximately 1 ml of mouse blood was diluted 1:1 with PBS with 10 mM EDTA pH 7.4 and a total volume of 1 ml of diluted blood was processed on each ^{Hb}CTC-Chip, using a Harvard Apparatus volumetric syringe pump set at 1.5 ml h^{-1}. After processing the blood, devices were washed with PBS followed by analyses as described below.

Human blood specimens obtained for CTC analysis was obtained after informed consent, per IRB protocol (05-300) at the Massachusetts General Hospital (MGH). A maximum of 20 ml of blood were obtained at any given blood draw, using EDTA vacutainers. Blood specimens were processed through the ^{Hb}CTC-Chip within 4 h of blood draw. Approximately 3 ml of blood was processed through the ^{Hb}CTC-Chip, using the pressure controlled apparatus, at a target flow rate of 1.5 ml h^{-1} as previously described[2].

CTC staining and enumeration. Following blood processing, captured cells on the ^{Hb}CTC-Chip were fixed with 4% paraformaldehyde and washed with PBS. The fixed cells were then permeabilized with 1% NP40 in PBS, blocked with either 5% donkey serum (mouse samples) or 3% goat serum/2% BSA (human samples), and then immunostained with the relevant primary antibodies. Primary antibodies for the mouse CTC analyses were rabbit anti-wide spectrum cytokeratin (1:50, Abcam ab9377), chicken anti-GFP (1:1,000, Abcam ab13970) and goat anti-mouse CD45 (1:500, R&D systems AF114). Primary antibodies for the human CTC analyses were rabbit anti-wide spectrum cytokeratin (1:50, Abcam ab9377), rabbit anti-EPCAM (1:500, Abcam ab71916), and mouse IgG1 anti-CD45 (1:1,000, BD 55480). Secondary immunofluorescent-tagged antibodies were used for signal amplification. For the mouse, these were donkey anti-rabbit Alexa Fluor 594 (1:500, Invitrogen A-21207), donkey anti-chicken Dylight 488 (1:500, Jackson ImmunoResearch 703-486-155) and donkey anti-goat Alexa Fluor 488 (1:500, Invitrogen A-11055). For human, the secondary antibodies were goat anti-rabbit Alexa Fluor 594 (1:500, Invitrogen A-11012) and goat anti-mouse IgG1 Alexa Fluor 488 (1:500, Invitrogen A-21121). Nuclei were then stained with DAPI and the devices were washed with PBS and stored at 4 °C. The devices were imaged under ×10 magnification using the BioView Ltd automated imaging system as well as an automated upright fluorescence microscope (Eclipse 90i, Nikon). Positive staining for cytokeratin or EPCAM, without CD45 staining, was required for scoring potential CTCs, which were then manually reviewed. Threshold and baseline signals were established using specimens from non-tumour-bearing mice or healthy human controls.

RNA extraction and purification from the ^{Hb}CTC-Chip and from tissues. RNA from CTC-enriched cell populations was isolated from the devices using a modified protocol for the Qiagen RNeasy MinElute kit (Qiagen). Briefly, 375 μl of the RLT buffer was pipetted into the ^{Hb}CTC-Chip with mixing between the inlet and outlet of the device. The resulting RLT lysate was then loaded onto the Qiagen RNeasy MinElute column and RNA was purified per protocol. RNA was eluted in a total of ~ 10 μl RNase-free water.

To isolate RNA from fresh frozen tissues, the material was pulverized with a sterile pestle in a microfuge tube on dry ice. RNA from cell lines and fresh frozen tumour and normal tissues were all processed in the same manner. RNA was extracted using the TRIzol reagent (Invitrogen) per manufacturer's specifications.

Single-molecule sequencing. Purified RNA was subjected to digital gene expression (DGE) sample prepping and analysed on the HeliScope Single Molecule Sequencer from Helicos BioSciences.

For mouse samples, a low quantity on surface RNA capture followed by complementary DNA synthesis and sequencing method was developed and used on RNA from EPCAM ^{Hb}CTC-Chip, IgG ^{Hb}CTC-Chip, primary tumour, and metastatic ascites. This method has been previously described[3]. Briefly, the purified RNA in RNase-free water was hybridized in 10-μl volume to Helicos poly(dT)-coated sequencing flow cell channels in 1× SSC, 0.05% SDS at 37 °C for 30 min. First-strand cDNA was synthesized with the SuperScript III first-strand cDNA synthesis kit (Invitrogen) using manufacturer's recommendations, except that no additional primers were added, and the incubation steps were modified as follows: 37 °C 15 min, and 55 °C 45 min. Subsequent to cDNA synthesis, hot water was passed through the channels to degrade and melt away the RNA strands. Guanine tailing was performed using terminal transferase, by adding 500 μM guanine in 20 μl volume in 1× TdT buffer, 2.5 mM $CoCl_2$ and 20 units terminal transferase per channel. The reaction took place at 37 °C for 30 min, followed by 3′ blocking with 100 μM ddGTP and ddATP under the same reaction conditions. The 18-nucleotide poly-C primers were hybridized at 50 nM in 1× SSC, 0.05% SDS at 55 °C for 30 min, followed by step-wise 'fill' steps with 500 μM cytosine and adenine nucleotides with 5 units Klenow fragment (NEB) in 1× NEB2 buffer and 20 μl reaction volume per channel. The lock step was then performed with virtual terminator guanine and thymidine nucleotide analogues. Single-molecule sequencing by synthesis was then initiated using standard procedures[12].

For human CTC samples, initial low quantity on surface methods did not have sufficient yield, so application of a low quantity RNA-seq method was used as previously described[13]. Briefly, the purified RNA in RNase-free water was converted to first-strand cDNA with the SuperScript III first-strand cDNA synthesis kit (Invitrogen) using random hexamers according to the manufacturer's recommendations. RNA was digested and single-stranded cDNA was purified using a combination of QIAquick nucleotide removal kit (Invitrogen) and ethanol precipitation with ammonium acetate and glycogen. Single-stranded cDNA was denatured and then a poly-A tail was added to the 3′ end using terminal transferase (New England Biolabs). Tailed cDNAs were then hybridized to the sequencing flow cell followed by 'fill and lock' and subjected to single-molecule sequencing by synthesis.

Bioinformatics. DGE: to compute digital gene expression (DGE) from the Helicos sequence data, we used the DGE pipeline of the HeliSphere 1.1.498.63 software (http://open.helicosbio.com) using the Human.Txome and Mouse.Txome (ftp://ftp.helicosbio.com/pub/distribution) reference files[14]. We ignored DGE output corresponding to mitochondrial and ribosomal RNA and Helicos control spike-ins.

DEGSeq: DEGseq comparisons of DGE profiles were run using the DEGexp function of version 1.0.5 of the Bioconductor (http://www.bioconductor.org) DEGseq package. The 'method' argument was set to MARS for the MA-plot-based method with Random Sampling[5] and the threshold for statistical significance was a Benjamini–Hochberg q-value (that is, FDR estimate) of 10%.

hGSEA: we performed hypergeometric gene set enrichment analysis (hGSEA) as follows. To determine whether gene set A was enriched in gene set B, we used the hypergeometric distribution to test whether the overlap of A and B was larger than would be expected by chance had A and B been drawn randomly from all the genes quantified by the DGE. To account for multiple hypotheses testing, we applied the method of Benjamini and Hochberg to the p-values so generated for each analysis, yielding false discover rate (FDR) estimates. The odds ratio (OR) is the odds of a gene being in A given it is in B divided by the odds of that gene being in A given it is not in B. Gene sets evaluated were obtained from MSigDB (Broad Institute).

QuantiGene ViewRNA *in-situ* hybridization on ^{Hb}CTC-Chip. Following blood processing, ^{Hb}CTC-Chips were fixed with 4% paraformaldehyde, washed with PBS and dehydrated with ethanol at increasing concentrations (50%, 70%, and 100%). ^{Hb}CTC-Chips were submerged in 100% ethanol at −20 °C for storage. Of note, mouse metastatic ascites fluid was also run on ^{Hb}CTC-Chips for RNA-ISH analysis. Before further analysis, cells were rehydrated using decreasing concentrations of ethanol (70% and 50%) and washed with PBS. ISH was performed using QuantiGene ViewRNA protocols. ^{Hb}CTC-Chip were permeabilized with Working Detergent Solution (Affymetrix), and digested with Protease (Affymetrix) at 1:2,000 dilution in PBS. The ^{Hb}CTC-Chips were hybridized for 3 h at 40 °C with custom-designed QuantiGene ViewRNA probes against genes of interest (mouse: *Wnt2*, *Krt8* and *Krt18*; human: *WNT2*, *KRT7*, *KRT8*, *KRT18*, *KRT19* and *KRT23*). Unbound probes were flushed out with Wash Buffer (Affymetrix) and ^{Hb}CTC-Chips were stored overnight at 4 °C in Storage Buffer (Affymetrix). The bound probes were amplified the following day through PreAmp (Affymetrix)

hybridization for 1 h at 40 °C, followed by Amp (Affymetrix) hybridization for 1 h at 40 °C. Label Probes (Affymetrix) targeting the individual probe types were added for 1 h at 40 °C. ^{Hb}CTC-Chips were stained with DAPI, and images were taken using a Nikon 90i microscope.

QuantiGene ViewRNA *in-situ* hybridization on FFPE tissue. Mouse tissue for FFPE was prepared per standard protocol by the MGH Clinical and Research Pathology Cores. ISH was performed using QuantiGene ViewRNA protocols. Five micrometre sections were cut, fixed in 10% formaldehyde (Fisher Scientific), deparaffinized, boiled in pre-treatment solution (Affymetrix) and digested with proteinase K (Affymetrix). Sections were hybridized for 3 h at 40 °C with custom-designed QuantiGene ViewRNA probes against *Wnt2* and the control gene *Ubc* (Affymetrix). Bound probes were then amplified per protocol from Affymetrix using PreAmp and Amp molecules. Multiple Label Probe oligonucleotides conjugated to alkaline phosphatase (LP-AP Type 1) are then added and Fast Red Substrate is used to produce signal (red dots, Cy3 fluorescence). For two colour assays, an LP-AP type 6 probe is used with Fast Blue substrate (blue dots, Cy5 fluorescence) followed by LP-AP type 1 probe with Fast Red Substrate (red dots, Cy3 fluorescence) to produce a dual colorimetric and fluorescent signal. *Wnt2* probes were used in type 1/Fast Red and pooled *Krt8* and *Krt18* were used in type 6/Fast Blue. Slides are then counterstained with haematoxylin. Serial sections were also subjected to haematoxylin and eosin staining per standard histology protocol to confirm the identity of cells in the region of RNA-ISH. Images were taken by a Nikon 90i scope with colour camera.

Constructs and viral infection. The mouse *Wnt2* open reading frame was cloned into pENTRI plasmid as a BamHI–XhoI fragment. pWPI plasmid carrying *Wnt2* open reading frame was generated by gateway cloning following the manufacturer's protocol (Invitrogen). Luciferase gene was expressed in pMSCV plasmid containing hygromycin selection marker. 7TFP construct (7xTcf-FFluc) used in TopFlash assay was from Addgene. pLKO shRNAs targeting the mouse *Fn1* and *Tak1* were from the RNAi consortium at the Broad Institute. Control shRNA constructs for Luciferase (SHC007, target sequences - GCGCGATAGCGCTAAT AATTT) and non-target (SHC016, target sequences - CGCTGAGTACTTCGAA ATGTC) were purchased from Sigma. Target sequences for *Fn1*: 5′-GCCTAGA AATACCTTTCTCTT-3′; for *Tak1*: B4 5′-AGGCAAAGCAACAGAGTGAAT-3′; B7 5′-TCTGAGAGGAAGGCTTTCATT-3′; B6 5′-CAGCCCTAGTGTCAGAA TGAT-3′; for human *TAK1*: 5′-GACACACATGACCAATAACAA-3′. Vesicular stomatitis virus glycoprotein-pseudotyped retroviruses or lentiviruses were generated using 293T cells as packaging cell lines following protocols described in the RNAi Consortium (Broad Institute). NB508 cells were infected with pWPI virus followed by GFP-positive cell sorting. pMSCV-luciferase was introduced into the cells and followed by 200 μg ml^{-1} hygromycin B selection. Cells transduced with pLKO shRNA virus were selected with 2 μg ml^{-1} puromycin. On-target plus siRNA constructs were purchased from Dharmacon and transfected into cells with Lipofectamine RNAiMax (Invitrogen). Target sequences for MYCN (L-058793-01): CCGGUGAACAAGCGAGAGA; UCGAAUUGGGCUACGGAGA; GAGGAU ACCUUGAGCGACU; UCUAACAACAAGGCGGUAA; ETV4 (L-048237-01): GGUCAGAGCUCCACUAUCC; UGAUCAAACAGGAGCGCAC; GAUACUU GGACCAGCGAGU; GAAAUGGGAGCUUGGGCGA; non-target (D-001810-10-20): UGGUUUACAUGUCGACUAA; UGGUUUACAUGUUGUGUGA; UGGUUUACAUGUUUUCUGA; UGGUUUACAUGUUUUCCUA.

Tumorigenicity and experimental metastasis assays. The animal protocol was approved by the MGH Subcommittee on Research Animal Care. Six-week-old Nu/Nu mice were anesthetized by isofluorane, 1×10^6 NB508 cells expressing vector, or WNT2 in 100 μl of PBS were injected subcutaneously on the right flank of the mice. Primary tumours, CTCs and lungs were sampled after 2 weeks. Six to 8-week old FVB mice were used for tail vein injection. NB508 cells (5×10^4) expressing luciferase and vector, or WNT2 in 100 μl of PBS were injected into lateral tail vein of the mice. Tumour formation in the lung was monitored weekly by bioluminescence using IVIS Lumina II (Caliper Life Science). Mice were euthanized after 6 weeks and lungs were sampled. Comparison between groups was performed using *t*-test.

Tumour sphere assay. Cells were plated as single-cell suspension in ultralow attachment 6-well or 96-well plates (Corning) and grown in RPMI medium (serum free) supplemented with 20 μl ml^{-1} B27 (Invitrogen), 20 ng ml^{-1} EGF and 20 ng ml^{-1} bFGF. Fresh media (1 ml or 30 μl) was added every 3 days. Tumour spheres were counted and photographed at day 10. Inhibitors were purchased from Sigma and used at 1 μM (for mouse cell lines) and 0.3, 1 and 3 μM (for human cell lines) (5Z-7-Oxozeaenol), 1.8 μM (IWR1), and 1.4 μM (Y-27632). For immunofluorescent staining, tumour spheres were deposited onto a microscope slide by cytospin (Shandon), fixed and stained. Comparison between groups was performed using *t*-test analysis (number differences) or Wilcoxon analysis (size differences).

Anoikis assay. Plates were coated with 10 ml of 20 mg ml^{-1} poly-HEMA (Sigma) and air-dried. Cells (2×10^6) were plated on the poly-HEMA-coated plates. Cells were collected at different time points and proteins extracted.

Antibodies and immunoblot analysis. Antibodies used were against β-catenin (BD), activated β-catenin (Millipore), Fibronectin, WNT2 (Sigma), β-tubulin (Santa Cruz), and cleaved Caspase-3, p-TAK1 (Cell Signaling), and TAK1 (Cell Signaling). Cells were lysed in RIPA lysis buffer (20 mM Tris, pH 8/150 mM NaCl/ 10 mM NaF/0.1% SDS/1% Nonidet P-40/1× protease inhibitor mixture (Roche)). For conditioned media, Vec-NB508 or Wnt2-NB508 cells were seeded at 1×10^6 in regular growth media (10% FBS) for 2 days in culture, and culture medium was replaced with low-serum media (0.2% FBS). After 24 h, conditioned medium was collected and cell debris was removed by centrifugation. Proteins were concentrated using Amico Ultracel 30k column (Millipore) and resuspended in RIPA buffer to detect the WNT2 protein by immunoblot. Lysates were run on an SDS/4–15% polyacrylamide gel (Bio-Rad) and transferred onto PVDF membranes (Invitrogen), and immunoblots were visualized with a Western Lightning Plus chemiluminescence kit (PerkinElmer). For phospho-TAK1 level, 1μg of protein lysate was immunoprecipitated with TAK1 antibody (Bethyl Laboratories) and blotted with p-TAK1 antibody (Cell Signaling).

SYTO60 assay. Cells were seeded in 96-well plates in standard growth media at 1,000 cells per well or serum-free media at 5,000 cells per well. Cells were fixed every day after plating and incubated with Syto60 red fluorescence nucleic acid staining (Invitrogen) at 1:8,000 dilution in 1× PBS for 30 min in the dark. Red fluorescence nucleic acid staining was measured using a plate reader (Molecular Probes).

Transwell migration and invasion assays. Cells (5×10^4) were seeded on 8-μm pore size Transwell filters (Corning) in regular growth media or low-serum media (0.2% FBS). Invasion chambers coated with Matrigel (Corning) were used for invasion assays with cells plated in regular growth media. Two days after seeding, cells on top of the chamber were stripped off with cotton tips and cells on the bottom of the chambers were fixed in 4% PFA and stained with crystal violet.

Real-time qRT–PCR. Total RNA (2 μg) was reverse transcribed using a cDNA synthesis kit (GE Healthcare). Real-time quantitative RT–PCR was performed using SYBR green in an ABI PRISM 7500 sequence detection system with 96-block module and automation accessory (Applied Biosystems). *Gapdh* was used as an internal control gene. The primers used are shown below: *Gapdh*: 5′-TGGTGAAGCAGGCATCTGAG-3′ and 5′-TGCTGTTGAAGTCGCAGGAG-3′. *Etv4*: 5′-CGACTCAGATGTCCCTGGAT-3′ and 5′-GGAATGGTCGAAGGGA TTTT-3′. *Mycn*: 5′-GATGAGGATGACGAGGAGGA-3′ and 5′-ACGCACAG TGATCGTGAAAG-3′. *Fn1*: 5′-GATAAATCAGCAGTGGGAACGG-3′ and 5′-CAAAGCAAGTCTCTTCAGGCTCAG-3′.

TOPflash assay. Vec or Wnt2-NB508 cells were infected with lentivirus carrying 7TFP construct and selected with puromycin. Cells were plated in 96-well plate at 5×10^3 cells per well. Small molecule reagents were added into media 1 day after plating and cells were imaged in the IVIS Lumina II (Caliper Life Science) the following day.

Low quantity high throughput qRT–PCR. RNA was isolated from adherent monolayer cells and tumour spheres cultured in non-adherent plate using Qiagen RNeasy kit (Qiagen). Reverse transcription was carried out using 200 ng of RNA with SuperScript III (Invitrogen) per manufacturer's protocol. Pre-amplification of cDNA was done with 1.25 μl of cDNA using pooled 48 TaqMan Assays at a final concentration of 0.2× for each assay. The pre-amplification PCR was performed at one cycle 95 °C for 10 min, 14 cycles at 95 °C for 15 s and then 60 °C for 4 min. After pre-amplification PCR, the product was diluted 1:5 with dH_2O and stored at −80 °C until needed. qPCR was carried out using the 48.48 dynamic array (Fluidigm Corporation) following the manufacturer's protocol.

12. Bowers, J. *et al.* Virtual terminator nucleotides for next-generation DNA sequencing. *Nature Methods* **6**, 593–595 (2009).
13. Ozsolak, F. *et al.* Digital transcriptome profiling from attomole-level RNA samples. *Genome Res.* **20**, 519–525 (2010).
14. Lipson, D. *et al.* Quantification of the yeast transcriptome by single-molecule sequencing. *Nature Biotechnol.* **27**, 652–658 (2009).

doi:10.1038/nature11183

Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion

Ravid Straussman[1], Teppei Morikawa[2], Kevin Shee[1], Michal Barzily-Rokni[1], Zhi Rong Qian[2], Jinyan Du[1], Ashli Davis[1], Margaret M. Mongare[1], Joshua Gould[1], Dennie T. Frederick[3], Zachary A. Cooper[3], Paul B. Chapman[4], David B. Solit[4,5], Antoni Ribas[6,7], Roger S. Lo[7,8], Keith T. Flaherty[3], Shuji Ogino[2,9], Jennifer A. Wargo[3] & Todd R. Golub[1,10,11,12]

Drug resistance presents a challenge to the treatment of cancer patients. Many studies have focused on cell-autonomous mechanisms of drug resistance. By contrast, we proposed that the tumour micro-environment confers innate resistance to therapy. Here we developed a co-culture system to systematically assay the ability of 23 stromal cell types to influence the innate resistance of 45 cancer cell lines to 35 anticancer drugs. We found that stroma-mediated resistance is common, particularly to targeted agents. We characterized further the stroma-mediated resistance of BRAF-mutant melanoma to RAF inhibitors because most patients with this type of cancer show some degree of innate resistance[1–4]. Proteomic analysis showed that stromal cell secretion of hepatocyte growth factor (HGF) resulted in activation of the HGF receptor MET, reactivation of the mitogen-activated protein kinase (MAPK) and phosphatidylinositol-3-OH kinase (PI(3)K)–AKT signalling pathways, and immediate resistance to RAF inhibition. Immunohistochemistry experiments confirmed stromal cell expression of HGF in patients with BRAF-mutant melanoma and showed a significant correlation between HGF expression by stromal cells and innate resistance to RAF inhibitor treatment. Dual inhibition of RAF and either HGF or MET resulted in reversal of drug resistance, suggesting RAF plus HGF or MET inhibitory combination therapy as a potential therapeutic strategy for BRAF-mutant melanoma. A similar resistance mechanism was uncovered in a subset of BRAF-mutant colorectal and glioblastoma cell lines. More generally, this study indicates that the systematic dissection of interactions between tumours and their micro-environment can uncover important mechanisms underlying drug resistance.

Oncoprotein-targeted drugs hold enormous promise for the future of cancer treatment. However, complete clinical responses are rare, suggesting that mechanisms exist to render a substantial proportion of tumour cells resistant to treatment. For example, melanomas harbouring a mutant *BRAF* gene encoding BRAF(V600E) (a mutant in which valine has been replaced with glutamic acid at residue 600) show marked responses to RAF inhibitors; however, these responses are almost always partial, and the tumours often recur within 6 months of treatment[1–4].

We proposed that innate drug resistance might be caused, at least in part, by factors secreted by the tumour micro-environment. The growth- and metastasis-promoting effects of the micro-environment have been well documented[5,6], but a possible role in drug resistance has been only partially explored[7–11]. To test the hypothesis that stromal cells might confer innate resistance on cancer cells, we developed a co-culture system whereby green fluorescent protein (GFP)-labelled tumour cells are co-cultured with stromal cells, and the ability of the stromal cells to modulate drug sensitivity is measured by monitoring the GFP levels over time (Supplementary Fig. 1). Forty-five GFP-labelled human cancer cell lines were cultured either alone or in combination with a panel of up to 23 human stromal cell lines in the presence of increasing doses of 35 widely used anticancer drugs (Supplementary Tables 1 and 2).

Our analysis of interactions between cancer cells, stromal cells and anticancer drugs (Supplementary Tables 3 and 4) showed a striking result—anticancer drugs that are capable of killing tumour cells are frequently rendered ineffective when the tumour cells are cultured in the presence of stromal cells (Fig. 1a). For example, certain dermal fibroblasts were able to confer complete resistance to the cytotoxic agent gemcitabine on colorectal and pancreatic cancer cell lines (Fig. 1b and Supplementary Fig. 2). Different stromal cells conferred resistance to RAF inhibitors on BRAF-mutant melanoma cell lines and to ERBB2 inhibitors on ERBB2-overexpressing breast cancer cell lines (Fig. 1c, d and Supplementary Figs 3 and 4). The stroma-mediated resistance phenomenon was particularly marked for targeted agents compared with conventional cytotoxic chemotherapy ($P < 0.001$; Supplementary Table 2). Overall, of the 23 targeted agents in the panel, there was evidence of micro-environment-mediated resistance to 15 (65%) (Supplementary Table 2 and Supplementary Information).

We next explored the mechanism of stroma-mediated innate resistance to the RAF inhibitor PLX4720 (an analogue of which, vemurafenib, was recently approved by the US Food and Drug Administration (FDA) for the treatment of BRAF-mutant melanoma). In a recent phase 3 clinical trial, 48% of patients with BRAF-mutant melanoma who were treated with vemurafenib had a confirmed response, but only 0.9% of patients had a complete response, indicating a high rate of innate resistance[2]. We tested 18 stromal cell lines for their ability to confer PLX4720 resistance on 7 BRAF(V600E) melanoma cell lines. Of these stromal cell lines, six fibroblast lines conferred resistance (Fig. 1c and Supplementary Fig. 3).

To determine whether the rescue effect was mediated by direct contact between fibroblasts and tumours or by the secretion of soluble factors, we tested the ability of fibroblast-conditioned growth medium to recapitulate the resistance effect. Fibroblast-conditioned medium was able to rescue BRAF-mutant melanoma cells from PLX4720, indicating that the rescue was due to a factor secreted by the fibroblasts (Fig. 2a). To identify the secreted factor with the rescue activity, we carried out an antibody-array-based analysis of 567 secreted factors (Supplementary Tables 5 and 6), comparing the conditioned medium from the 6 stromal cell lines with rescue activity to that from 12 stromal

[1]The Eli and Edythe L. Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02115, USA. [3]Division of Surgical Oncology, Medical Oncology and Dermatology, Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114, USA. [4]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [5]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [6]Division of Hematology and Oncology, Department of Medicine, Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California 90095, USA. [7]Department of Molecular and Medical Pharmacology, Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California 90095, USA. [8]Division of Dermatology, Department of Medicine, Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California 90095, USA. [9]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [10]Department of Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. [11]Harvard Medical School, Boston, Massachusetts 02115, USA. [12]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

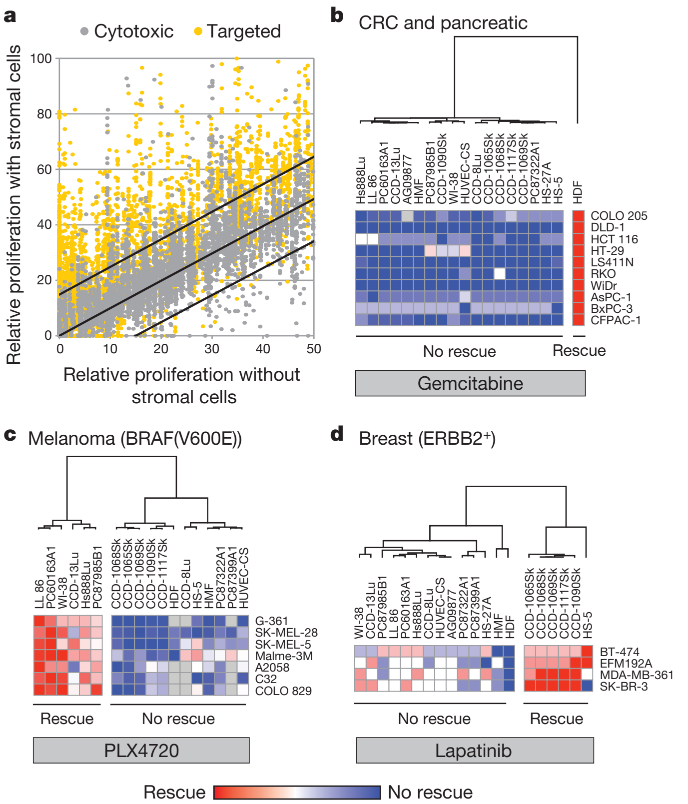


Figure 1 | The effect of stromal cells on the chemoresistance of cancer cell lines. a, Forty-five GFP-labelled cancer cell lines were treated with 35 cytotoxic or targeted anticancer drugs, either alone or in co-culture with a panel of up to 23 stromal cell lines and primary cells. The relative proliferation was calculated by normalizing the number of cells (as determined by GFP fluorescence) after 7 days of treatment to the number of cells (GFP) in the dimethylsulphoxide (DMSO) control wells. The *x* axis shows relative proliferation in the absence of stromal cells, whereas the *y* axis shows relative proliferation in the presence of stromal cells. The *y* axis values were also normalized to the effect that each stromal cell type had on cancer cell proliferation when no drug is present (to distinguish true rescue from stromal effects on proliferation). The central diagonal line represents the expected result when stromal cells do not confer resistance. The upper and lower diagonal lines represent one s.d. from the central diagonal line. **b**, Hierarchical clustering of stromal cells (horizontal) according to their ability to rescue colorectal (CRC) and pancreatic cancer cell lines (vertical) from 0.1 μM gemcitabine. **c**, Hierarchical clustering of stromal cells according to their ability to rescue melanoma cancer cell lines with the BRAF(V600E) mutation from 2 μM PLX4720. **d**, Hierarchical clustering of stromal cells according to their ability to rescue *ERBB2*-amplified breast cancer cell lines from 2 μM lapatinib. See Supplementary Figs 2–4 for details.

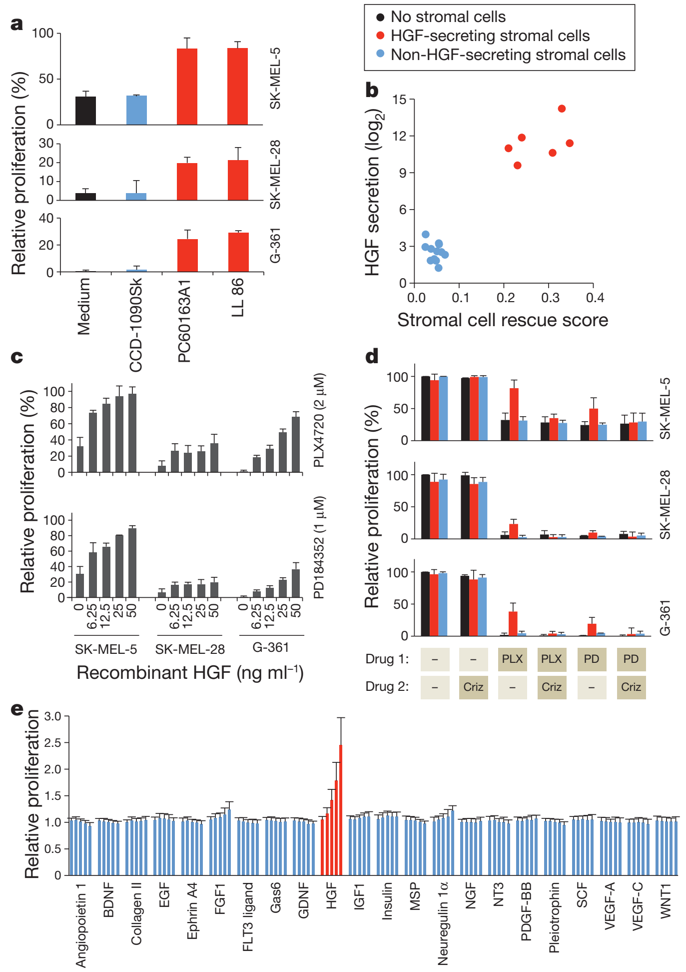


Figure 2 | HGF rescues melanoma cancer cell lines from RAF and MEK inhibitors. a, Three melanoma cell lines (vertical) were co-cultured with conditioned medium from three fibroblast cell lines (horizontal) or with fresh medium and then treated with 2 μM PLX4720. Proliferation was quantified after 7 days and compared with that of non-treated cells. The bars represent the mean ± s.e.m. between replicates ($n = 3$). **b**, The HGF secretion level of 18 stromal cell lines measured by a protein cytokine array (Supplementary Table 5) is plotted against the ability of each stromal cell line to rescue the BRAF(V600E) melanoma cell lines from PLX4720 (Supplementary Fig. 5). **c**, The effect of HGF (6.25–50 ng ml^{-1}) on the proliferation of melanoma cell lines under PLX4720 or PD184352 treatment. The bars represent the mean ± s.e.m. between replicates ($n = 3$). **d**, Drug resistance manifests only in the presence of HGF-secreting stromal cells and is reversed by a MET inhibitor. Melanoma cell lines were co-cultured with nine stromal cell lines, representing HGF-secreting and non-secreting stromal cells, and treated with 2 μM PLX4720 (PLX) or 1 μM PD184352 (PD) with or without 0.2 μM crizotinib (Criz). Proliferation was quantified after 7 days and was normalized to that of non-treated cells. The results were averaged across four stromal cell lines that secrete HGF and five that do not. Non-averaged results are presented in Supplementary Fig. 11. The bars represent the mean ± s.e.m. between replicates ($n = 3$). **e**, Twenty-two cytokines, each at five concentrations (Supplementary Table 8), were added to six melanoma cell lines that were then treated with 2 μM PLX4720, 1 μM PD184352 or DMSO control. Proliferation was quantified after 7 days and was normalized to no-cytokine controls. The results shown are averaged for all cell lines and both drugs. The bars represent the mean ± s.e.m. between replicates ($n = 3$).

cell lines without rescue activity. The factor that correlated best with PLX4720 resistance was HGF (Fig. 2b and Supplementary Figs 5 and 6), a well-characterized growth factor whose secretion by mesenchymal cells induces activation of the receptor tyrosine kinase MET. Although MET has been reported to be overexpressed in melanoma[12,13] and to contribute to melanoma progression[12], it has not previously been implicated in RAF-inhibitor resistance. However, there have been recent reports that MET activation has a potential role in the development of resistance to gefitinib, an epidermal growth factor (EGF) receptor inhibitor, in non-small-cell lung cancer[10,14].

We next used immunohistochemistry to examine HGF expression in 34 biopsy samples derived from patients with BRAF(V600E) melanoma that were taken immediately before treatment with a RAF inhibitor (or a combination of a RAF inhibitor and a MAPK kinase 1 (MEK1) or MEK2 inhibitor). HGF was detected in the tumour-associated stromal cells in 23 of the 34 patients (68%) (Fig. 3a, b and Supplementary Table 7), and phosphorylated MET (pMET) immunofluorescence studies accordingly showed MET phosphorylation (activation) in patient samples (Supplementary Fig. 7).

Our *in vitro* studies predict that the presence of stromal HGF should be associated with innate resistance. Indeed, patients whose stromal cells secreted HGF had a significantly poorer response to treatment than those lacking HGF expression ($P < 0.05$; Fig. 3c). Interestingly, only 1 of the 34 patients had a durable complete response (ongoing at

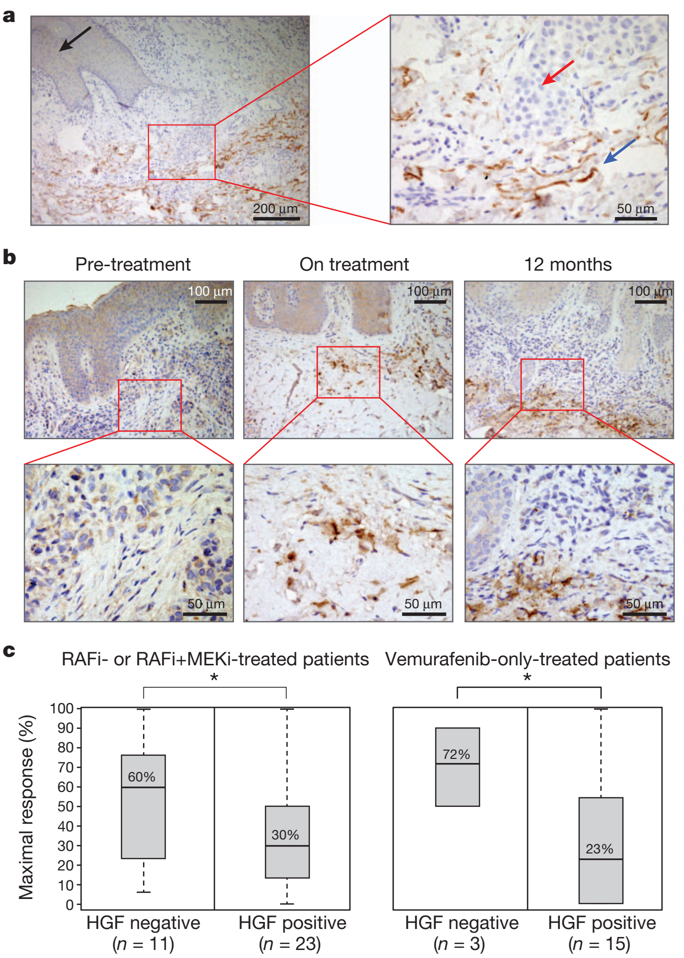


Figure 3 | HGF is present in the stromal cells of melanoma and correlates with a poor response to therapy. **a**, A pre-treatment melanoma section from patient 32 was analysed for HGF expression by immunohistochemistry: black arrow, normal epidermis; red arrow, tumour cells; and blue arrow, HGF-expressing stroma (brown). A high-magnification image of the red boxed area is shown on the right. **b**, Melanoma sections from patient 23 were analysed for HGF expression using immunohistochemistry. The on-treatment biopsy sample was obtained 2 weeks after the initiation of treatment with the BRAF inhibitor vemurafenib (PLX4032) and 1 month after the pre-treatment biopsy was obtained. A third biopsy sample was obtained 12 months after the initiation of treatment while the patient was progressing under treatment. High-magnification images of the red boxed areas are shown in the lower images. **c**, Maximal response to treatment with RAF inhibitor (RAFi), a RAFi and MEK inhibitor (MEKi) combination or the RAF inhibitor vemurafenib in patients with BRAF(V600E)-carrying melanoma that expressed or did not express stromal HGF, as determined by immunohistochemistry. Patients with stromal HGF had a significantly poorer response to treatment than those lacking HGF expression. *$P < 0.05$ by a two-sample t-test assuming equal variance. In the boxplots, the boxes cover the interquartile range (IQR) between the first and the third quartiles (Q1 and Q3). The whiskers are drawn at the extreme value that is no more than Q3 + 1.5 × IQR, and no less than Q1 − 1.5 × IQR. The median values for each group are depicted above the median line.

14 months after therapy), and this patient lacked HGF expression (Supplementary Table 7). On-treatment biopsy samples that were taken 2 weeks after treatment initiation were also available for 10 patients. For 5 of those (50%), stromal HGF expression was higher than before treatment (Fig. 3b and Supplementary Table 7). Whether this increase is attributable to the recruitment of HGF-secreting fibroblasts to the tumour or to the upregulation of HGF expression in the existing fibroblasts remains to be determined. It is notable that both normal skin and benign naevi showed stromal HGF expression (Supplementary Fig. 8). Our results thus support the clinical relevance of HGF-mediated resistance to BRAF inhibitors. Importantly, a similar inverse association between plasma HGF levels and response to RAF inhibitor treatment in patients with BRAF-mutant melanoma has been found in another study[15].

To establish HGF as the cause of drug resistance and not simply a biomarker for resistance, we tested the ability of recombinant HGF to induce resistance, as well as the ability of HGF-neutralizing antibodies or the MET-inhibitory small-molecule crizotinib to block fibroblast-induced PLX4720 resistance. These experiments indicated that HGF is both necessary and sufficient to confer the resistance phenotype (Fig. 2c, d and Supplementary Figs 9–11). Consistent with this observation, the extent to which 20 BRAF-mutant melanoma cell lines could be rescued by HGF was highly correlated with the MET expression level (Supplementary Fig. 12).

Although our stromal-cell-profiling studies pointed to the HGF–MET axis as the most relevant in mediating PLX4720 resistance, it is conceivable that other ligands for receptor tyrosine kinases (RTKs) might similarly confer resistance. To test this possibility, we tested the ability of 22 well-characterized RTK ligands to rescue BRAF-mutant melanoma cells from either PLX4720 or the MEK inhibitor PD184352. Surprisingly, despite many RTKs being expressed and activated by their cognate ligands, HGF was the only ligand that conferred substantial resistance to RAF or MEK inhibition (Fig. 2e, Supplementary Figs 13–15 and Supplementary Table 8).

We next sought to clarify the precise mechanism by which the HGF–MET axis is uniquely capable of inducing primary resistance to PLX4720. MET is known to activate both the MAPK pathway (the MEK–extracellular signal-regulated kinase (ERK) pathway) and the PI(3)K pathway (the PI(3)K–AKT pathway) (Supplementary Fig. 16), and both pathways have been suspected of being involved in acquired resistance to BRAF inhibitors[16–18]. We used western blot analysis to assess the ERK and AKT activation status of a panel of 7 BRAF-mutant melanoma cell lines treated with a RAF inhibitor together with various RTK ligands. HGF treatment led to the sustained activation of both ERK and AKT, but this dual activation was not observed for any of the other RTK ligands in any of the melanoma lines (Fig. 4a and Supplementary Fig. 17).

It should be noted that although EGF, fibroblast growth factor 1 (FGF1) and the BB form of platelet-derived growth factor (PDGF-BB) were able to reactivate ERK in most cell lines, the levels of phosphorylated ERK (pERK) were modest compared with those of HGF-treated cells. Moreover, these ligands failed to activate AKT. Similarly, treatment with insulin or insulin-like growth factor 1 (IGF1) led to a transient increase in phosphorylated AKT (pAKT) but did not activate ERK (Fig. 4b and Supplementary Fig. 18).

HGF was thus unique in its ability to induce the sustained activation of both ERK and AKT (Fig. 4c and Supplementary Figs 19 and 20). Importantly, we found that the HGF-mediated activation of ERK was more profound under BRAF inhibition than under MEK inhibition (Fig. 4c and Supplementary Fig. 19). This might be best explained by the finding that in the presence of BRAF inhibitors, MET can reactivate MEK through RAF1 (also known as CRAF), thus bypassing BRAF; however, this pathway is not possible under conditions of direct MEK inhibition (Fig. 4c and Supplementary Fig. 16).

Our model thus predicts that both the MAPK pathway and the PI(3)K–AKT pathway contribute to the primary resistance induced by HGF-secreting stromal cells. In agreement with this model, we have found that HGF-induced resistance is greater under BRAF inhibition than MEK inhibition (Fig. 2d), that combination treatment with a BRAF inhibitor and a MEK inhibitor is not sufficient to eliminate HGF-induced resistance (as this combination does not silence AKT (Supplementary Fig. 21)), and that combination treatment with a MEK inhibitor and an AKT inhibitor suppresses the majority of HGF-induced drug resistance (Supplementary Fig. 21).

Our discovery of HGF-mediated innate resistance to BRAF inhibitors should be distinguished from recent reports that propose the dysregulation of IGF1, PDGF, COT, BRAF or MEK as mechanisms of resistance

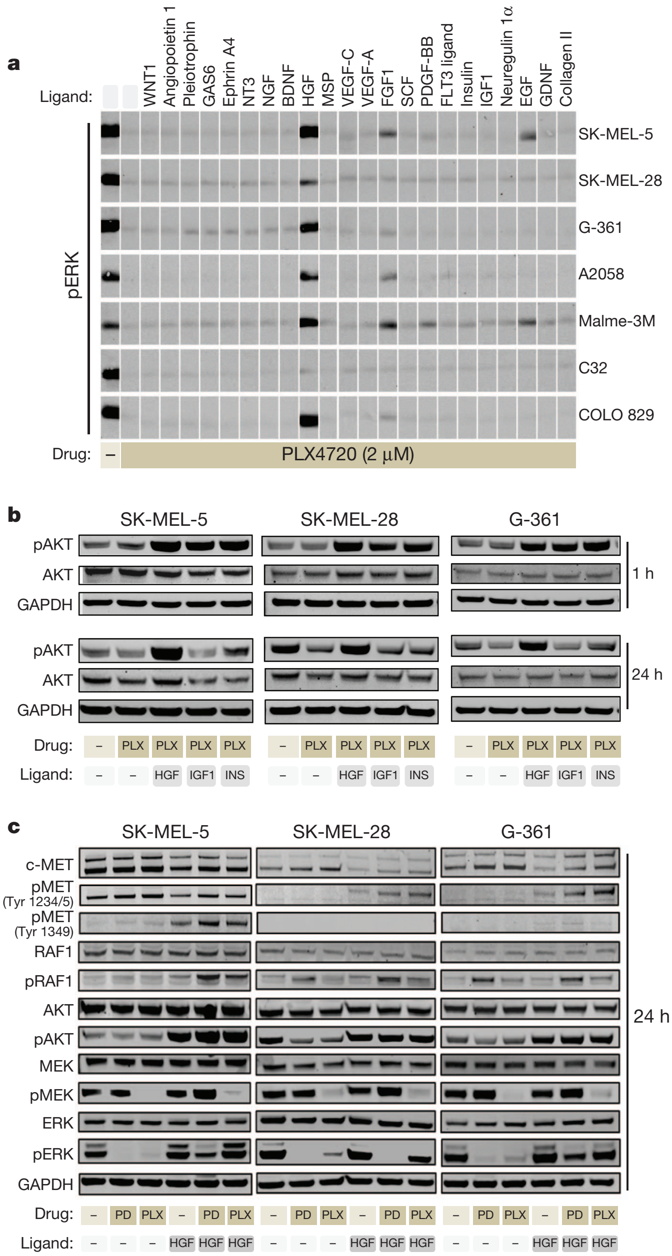


Figure 4 | Characterizing the molecular mechanism of HGF-induced primary resistance. **a**, Activation of ERK by cytokines. Levels of pERK (Thr 202/Tyr 204) in melanoma cell lines were assayed by western blot analysis 1 h after treatment with medium (–) or with each of 22 cytokines in the presence of PLX4720 or DMSO control. **b**, The activation of AKT in melanoma cell lines by HGF, IGF1 (IGF) or insulin (INS). The levels of pAKT (Ser 473) were assayed by western blot analysis 1 h and 24 h after treatment with HGF, IGF1 or insulin in the presence of 2 μM PLX4720 (PLX). **c**, The effect of 25 ng ml^{-1} HGF on melanoma cell lines treated with 2 μM PLX4720 or 2 μM PD184352 (PD). MAPK and PI(3)K–AKT pathway activation was assessed by western blot analysis after 24 h of treatment. GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

to RAF inhibitors[18–22]. In these reports, the emergence of late (acquired) drug resistance was studied (for example, following exposure to a drug for many months), whereas we found that HGF-secreting stromal cells confer immediate (innate) resistance to RAF inhibitors. For example, the p61 splice variant of BRAF(V600E) that was recently shown to confer resistance to RAF inhibitors[21] was not observed in tumours before RAF inhibitor treatment, implicating this splice variant as a mechanism of acquired rather than innate resistance. Whether HGF has a role in acquired resistance as well remains to be determined.

Activation of the EGF receptor was recently shown to drive the resistance of some BRAF(V600E)-carrying colorectal cancer cell lines to RAF inhibitors[23,24]. To explore a possible role for MET activation in BRAF-mutant non-melanoma cancers, we tested seven BRAF-mutant non-melanoma cell lines (five colorectal cancer lines and two glioblastoma lines) and found that all seven had evidence of MET phosphorylation (Supplementary Fig. 22). Although stromal HGF expression is less common in colorectal cancer than melanoma (Supplementary Fig. 8a), MET overexpression and HGF autocrine secretion have been documented in colorectal cancer[25–27]. Indeed, we identified two BRAF-mutant non-melanoma cell lines that secreted HGF: one colorectal cell line (RKO), and one glioblastoma cell line (KG-1-C) (Supplementary Fig. 6). In these two cell lines, combined RAF and MET (but not EGF receptor) inhibition resulted in a clear synergistic effect (Supplementary Figs 22 and 23). The synergy between RAF and MET inhibitors was more variable among non-HGF-secreting BRAF-mutant cell lines (Supplementary Fig. 22). As predicted from our proposed mechanism of resistance, monotherapy with RAF or MEK inhibitors had no effect on pAKT levels and caused little inhibition of pERK in HGF-secreting cell lines. However, dual inhibition of BRAF and MET resulted in significant inhibition of both pERK and pAKT (Supplementary Fig. 24). The extent to which autocrine or micro-environment-mediated MET activation explains the failure of BRAF-mutant non-melanoma tumours to respond to BRAF inhibition deserves further investigation.

The findings reported here potentially have immediate clinical implications. Several small-molecule or antibody inhibitors of HGF or MET are in clinical development or have been approved by the FDA for other indications. Given the tolerability of those agents and the similar tolerability of RAF inhibitors, combination clinical trials in BRAF-mutant melanoma, colorectal cancer and possibly other tumour types should be considered.

It should be noted that the stroma-derived, HGF-mediated RAF inhibitor resistance mechanism detailed here was but one of many such stroma-mediated drug resistance interactions uncovered in our initial screen (Fig. 1a). Our findings point to the micro-environment as an important but understudied source of anticancer drug resistance. Moreover, the results suggest that such resistance mechanisms can be uncovered through the systematic dissection of interactions between tumours and their micro-environment. Future studies should therefore seek to identify such resistance mechanisms for all of the drugs that are in development, potentially leading to mechanism-based combination regimens such as the RAF- and MET-inhibitor combination proposed here.

METHODS SUMMARY

Stroma-mediated chemoresistance co-culture screen. On day 0, stromal cells (1,700 cells in 20 μl per well) were plated in 384 clear-bottom plates (Corning, catalogue number 3712), together with GFP-labelled cancer cells (1,700 cells in 20 μl per well). On day 1, the cells were treated with 10 μl 5× drug using the CyBi-Well Vario 384/25 simultaneous pipettor (CyBio). On day 4, the medium in all of the wells was replaced with fresh medium, and fresh drug was added to all of the wells containing melanoma cell lines (all other cancers were treated on day 1 only). GFP fluorescence was read on days 1, 4 and 7, using a SpectraMax M5e Microplate Reader (Molecular Devices). A fluorescence microscope with high-throughput screening capabilities (Axio Observer.Z1, Zeiss) was used to document bright-field and GFP images on day 7. All screens were carried out in quadruplicate. See the Supplementary Information for a complete description of all other methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 November 2011; accepted 4 May 2012.
Published online 4 July 2012.

1. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467,** 596–599 (2010).

2. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364,** 2507–2516 (2011).
3. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363,** 809–819 (2010).
4. Sosman, J. A. *et al.* Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N. Engl. J. Med.* **366,** 707–714 (2012).
5. Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. *Nature Rev. Cancer* **9,** 239–252 (2009).
6. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nature Rev. Cancer* **6,** 392–401 (2006).
7. McMillin, D. W. *et al.* Tumor cell-specific bioluminescence platform to identify stroma-induced changes to anticancer drug activity. *Nature Med.* **16,** 483–489 (2010).
8. Shekhar, M. P., Santner, S., Carolin, K. A. & Tait, L. Direct involvement of breast tumor fibroblasts in the modulation of tamoxifen sensitivity. *Am. J. Pathol.* **170,** 1546–1560 (2007).
9. Teicher, B. A. *et al.* Tumor resistance to alkylating agents conferred by mechanisms operative only *in vivo*. *Science* **247,** 1457–1461 (1990).
10. Wang, W. *et al.* Crosstalk to stromal fibroblasts induces resistance of lung cancer to epidermal growth factor receptor tyrosine kinase inhibitors. *Clin. Cancer Res.* **15,** 6630–6638 (2009).
11. Williams, R. T., Roussel, M. F. & Sherr, C. J. *Arf* gene loss enhances oncogenicity and limits imatinib response in mouse models of *Bcr–Abl*-induced acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **103,** 6688–6693 (2006).
12. Puri, N. *et al.* c-Met is a potentially new therapeutic target for treatment of human melanoma. *Clin. Cancer Res.* **13,** 2246–2253 (2007).
13. Lee, Y. J. *et al.* Expression of the c-Met proteins in malignant skin cancers. *Ann. Dermatol.* **23,** 33–38 (2011).
14. Engelman, J. A. *et al. MET* amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316,** 1039–1043 (2007).
15. Wilson, T. R. *et al.* Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* http://dx.doi.org/10.1038/nature11249 (2012).
16. Jiang, C. C. *et al.* MEK-independent survival of B-RAF^{V600E} melanoma cells selected for resistance to apoptosis induced by the RAF inhibitor PLX4720. *Clin. Cancer Res.* **17,** 721–730 (2011).
17. Shao, Y. & Aplin, A. E. Akt3-mediated resistance to apoptosis in B-RAF-targeted melanoma cells. *Cancer Res.* **70,** 6670–6681 (2010).
18. Villanueva, J. *et al.* Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer Cell* **18,** 683–695 (2010).
19. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468,** 968–972 (2010).
20. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468,** 973–977 (2010).
21. Poulikakos, P. I. *et al.* RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature* **480,** 387–390 (2011).
22. Wang, H. *et al.* Identification of the MEK1(F129L) activating mutation as a potential mechanism of acquired resistance to MEK inhibition in human cancers carrying the *B-Raf*V600E mutation. *Cancer Res.* **71,** 5535–5545 (2011).
23. Corcoran, R. B. *et al.* EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* **2,** 227–235 (2012).
24. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483,** 100–103 (2012).
25. Kammula, U. S. *et al.* Molecular co-expression of the c-Met oncogene and hepatocyte growth factor in primary colon cancer predicts tumor stage and clinical outcome. *Cancer Lett.* **248,** 219–228 (2007).
26. Di Renzo, M. F. *et al.* Overexpression and amplification of the met/HGF receptor gene during the progression of colorectal cancer. *Clin. Cancer Res.* **1,** 147–154 (1995).
27. Liu, C., Park, M. & Tsao, M. S. Overexpression of c-met proto-oncogene but not epidermal growth factor receptor or c-erbB-2 in primary human colorectal carcinomas. *Oncogene* **7,** 181–185 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all members of the Golub laboratory for discussions. We thank C. M. Johannessen, I. Rabinowitch, N. Shoresh and A. Goren for critical reading of the manuscript, and L. Gaffney for expert assistance with the graphics. This work was supported by the Howard Hughes Medical Institute, National Cancer Institute grants P50CA093683 and U54CA112962 (T.R.G.) and a Melanoma Research Alliance Team Science Award.

Author Contributions R.S. and T.R.G. conceived and designed the experiments. R.S. performed the primary cancer–stroma–drugs screen with help from K.S., M.B.-R. and A.D. R.S. carried out the protein array experiments. R.S., K.S., M.B.-R., A.D. and M.M.M. carried out the secondary screens, western blot analysis and enzyme-linked immunosorbent assays (ELISAs). J.D. performed tyrosine kinase phosphorylation profiling. Clinical samples and clinical data were collected by J.A.W., K.T.F., D.T.F., P.B.C., D.B.S., A.R. and R.S.L. The immunohistochemistry experiments were carried out and analysed by S.O., T.M. and Z.R.Q. The immunofluorescence experiments were carried out by J.A.W. and Z.A.C. R.S. and T.R.G. produced the text and figures, including the Supplementary Information. K.S. helped to produce some of the text and figures. All authors discussed the results and contributed to the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.R.G. (golub@broadinstitute.org).

METHODS

Stroma-mediated chemoresistance co-culture screen. On day 0, stromal cells (1,700 cells in 20 μl per well) were plated in 384 clear-bottom plates (Corning, catalogue number 3712), together with GFP-labelled cancer cells (1,700 cells in 20 μl per well). On day 1, the cells were treated with 10 μl 5× drug using the CyBi-Well Vario 384/25 simultaneous pipettor (CyBio). On day 4, the medium in all of the wells was replaced with fresh medium, and fresh drug was added to all of the wells containing melanoma cell lines (all other cancers were treated on day 1 only). GFP fluorescence was read on days 1, 4 and 7, using a SpectraMax M5e Microplate Reader (Molecular Devices). A fluorescence microscope with high-throughput screening capabilities (Axio Observer.Z1, Zeiss) was used to document bright-field and GFP images on day 7. All screens were carried out in quadruplicate.

Cell lines and reagents. The sources of all used cell lines are listed in Supplementary Table 1. All cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen, 10569-010) with 10% FBS and 1× Pen-Strep-Glutamine (Invitrogen, 15140-122). Cancer cell lines were lentivirally transduced using the pLex_TRC206 plasmid. The sources of all used drugs are listed in Supplementary Table 2. Antibodies specific for MET (3148), pMET (3077 and 3133), pRAF1 (9427), pERK (4370), AKT (2920), pAKT (4060), MEK (4694), pMEK (9154) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (2118) were purchased from Cell Signaling Technology. Antibody specific for RAF1 (ab656) was purchased from Abcam. Antibody specific for ERK (sc-135900) was purchased from Santa Cruz Biotechnology. Anti-rabbit (926-32211) and anti-mouse (926-32220) secondary antibodies were purchased from LI-COR. The following cytokines were purchased from R&D Systems: angiopoietin 1 (923-AN-025), BDNF (248-BD-005), EGF (236-EG-200), ephrin A4 (369-EA-200), FGF1 (231-BC-025), FLT3 ligand (308-FK-025), GAS6 (885-GS-050), GDNF (212-GD-010), IGF1 (291-G1-050), MSP (352-MS-010), neuregulin 1α (5898-NR-050), NGF (256-GF-100), NT3 (267-N3-005), PDGF-BB (220-BB-010), pleiotrophin (252-PL-050), VEGF-A (293-VE-010) and VEGF-C (2179-VC-025). HGF (228-10702-2) was purchased from RayBiotech. Insulin (I9278) was purchased from Sigma. Stem cell factor (569600-10UG) was purchased from EMD. Type II collagen (ab7534) and WNT1 (ab84080) were purchased from Abcam. Skin tissue microarrays (TMAs) of normal skin, naevi and melanomas (ME1004a and ME803a) were purchased from US Biomax. Colorectal cancer TMAs were prepared as described previously[28].

Clinical samples. Patients with metastatic melanoma with the BRAF(V600E) mutation (confirmed by genotyping) were enrolled in clinical trials for treatment with a BRAF inhibitor or a combination of a RAF inhibitor and a MEK inhibitor (Supplementary Table 7), and the patients gave consent for tissue acquisition as per an IRB-approved protocol. Tumours were biopsied before treatment (day 0), at 10–14 days during treatment and at the time of progression. Formalin-fixed tissue was analysed (by haematoxylin and eosin staining) to confirm that a viable tumour was present. The tumour responses were determined by the investigators according to the Response Evaluation Criteria in Solid Tumours (RECIST).

Analysis of co-culture screen data. GFP fluorescence readings were measured for each well on day 7. Background fluorescence was removed from the readings by subtracting the readings for the same wells on day 1. Quadruplicates were then averaged. The drug effect on each cancer cell line in the presence or absence of stromal cells was calculated by normalizing the number of cells after 7 days of treatment (as measured by GFP fluorescence) to the number of cells (GFP) present in the dimethylsulphoxide (DMSO) control wells. The drug effect in the presence of stromal cells was normalized further to the effect that each stromal cell type had no cancer cell proliferation when no drug was present (see the 'without stroma' and 'with stroma' columns in Supplementary Table 3). The 'rescue score' was calculated by subtracting the 'without stroma' drug effect from the 'with stroma' drug effect.

Assigning a 'rescue by stroma' score to all screened drugs. See Supplementary Table 2 for further details. Only cases in which a drug slowed the proliferation of a cancer cell line to <30% were analysed (Supplementary Table 4). Rescue was counted as positive if the rescue score was >0.3. Drugs whose activity was rescued by stromal cells in at least three cancer cell lines and in >40% of all cancer cell lines screened with this drug obtained the maximum score: ++. Drugs whose activity was rescued by stromal cells in three cancer cell lines but in only 20–40% of cancer cell lines screened with this drug scored +, as did drugs whose activity was rescued in only 1 or 2 cell lines and in >40% of the cancer cell lines screened with this drug.

Antibody arrays. Soluble proteins in the medium of the stromal cell lines were measured using the Human Cytokine Array G4000 (RayBio, AAH-CYT-G4000-8) and a Biotin Label-Based Human Antibody Array (RayBio, AAH-BLG-1-4), according to the recommended protocols. These arrays can detect 274 and 507 proteins, respectively. Stromal cells were plated 3 days before the experiment in DMEM containing 10% FBS and were 75–90% confluent when the media were collected and filtered. Medium containing 10% FBS was also hybridized to the arrays and used later for normalization. Ten technical and biological replicates were carried out, and both showed a very high correlation (correlation coefficient >0.9) (data not shown). Hybridization was carried out overnight at 4 °C. All slides were scanned using a GenePix 4000B Microarray Scanner (Axon) and analysed using the software GenePix Pro 6.0. The F532 median − B532 score was used and averaged across triplicates on each array. The results were then normalized using internal controls, and the values for cytokines in clear medium containing 10% FBS were subtracted. All results are available in Supplementary Tables 5 and 6.

Stroma-averaged melanoma rescue scores. The averaged melanoma rescue effect of each stromal cell line was calculated by averaging the rescue scores of this cell line (Supplementary Table 3) across all melanoma cell lines and all PLX4720 concentrations. Only instances in which the treatment caused a drop in proliferation to <0.3 when no stromal cells were present were included in this calculation.

The effect of pre-conditioned medium (PCM). PCM was prepared by filtering the medium from 80–90% confluent 15-cm plates that had been plated 3 days earlier and then diluting this medium 1:1 with fresh medium. Experiments were performed according to the previously described co-culture experiment protocol except for the following changes. First, on day 0, 384-well plates were seeded with 20 μl per well of PCM instead of 20 μl stromal cells. Second, on day 1, the medium from all of the wells was changed to fresh PCM. Third, on day 4, the medium was changed to fresh PCM instead of fresh medium before retreating the cells.

Hierarchical clustering. Unsupervised hierarchical clustering of stromal cell lines according to their ability to rescue melanoma cancer cell lines from 2 μM PLX4720 (Supplementary Table 3) was carried out using GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/). The Euclidean distance metric was used.

HGF enzyme-linked immunosorbent assay (ELISA). Cells were plated 3 days before the experiment in DMEM containing 10% FBS and were 75–90% confluent when medium was collected and filtered. An HGF ELISA was performed using a Human HGF ELISA Kit (RayBio, ELH-HGF-001) according to the kit's instructions. The medium was diluted 1:1 with diluent B before it was added to the assay microplate. For the standard HGF curve, we used the same HGF that was used for all of the other experiments (228-10702-2) and not the HGF that was supplied with the kit. The absorbance was read at 450 nm using a SpectraMax M5e Microplate Reader.

Neutralizing HGF with anti-HGF antibodies. Co-culture experiments were performed as described above except for the addition of neutralizing anti-HGF antibodies (R&D Systems, MAB294) on day 0 and after the medium was changed on day 4.

Western blot analysis and quantification. Cells were plated 1 day before treatment in a 6-well plate at 5×10^5 cells per well, and were treated the next day. At the designated time points, cells were lysed with lysis buffer containing 50 mM Tris, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 1 mg ml^{-1} NaF, and one pellet per 10 ml each of PhosSTOP Phosphatase Inhibitor (Roche, 04906837001) and Complete Mini Protease Inhibitor (Roche). Protein concentrations were determined with a DC Protein Assay Kit II (Bio-Rad). Samples were mixed with 4× protein sample loading buffer (LI-COR, 928-40004) and NuPAGE Sample Reducing Agent (Invitrogen, NP0009), and proteins were separated on a 4–12% Bis-Tris gel (NuPAGE, WG1402BOX) at 120 V. Proteins were transferred onto membranes using Program 4 on the iBlot Gel Transfer Device (Invitrogen, IB1001). Western blot analysis was performed according to the antibody manufacturer's specifications. Near-infrared fluorescence was detected with the Odyssey Infrared Imaging System (LI-COR), and signal intensity was quantified with Odyssey Application Software (LI-COR). All values were first normalized to background intensity and then to a GAPDH loading control.

High-throughput western blot analysis. High-throughput western blot analysis experiments (Fig. 4a and Supplementary Fig. 17) were performed as described above, except for the following changes. First, cells were seeded in a 96-well plate at 5×10^4 cells per well. Second, samples were mixed with E-PAGE 4× Loading Buffer (Invitrogen, EPBUF-01), and proteins were separated on 6% E-PAGE 96-well gels (Invitrogen, EP09606). For the transfer to membranes, Program 3 of the iBlot Gel Transfer Device was used.

Rescue of melanoma cell lines by cytokines. One day before treatment, cancer cells were seeded in black 384-well plates (Corning, 3712) at a concentration of 2,500 cells per well. On day 1, all 22 ligands were added at 5 different concentrations to 6 melanoma cell lines treated with PLX4720, PD184352 or DMSO control. On day 4, the medium was changed to fresh medium, and the cells were retreated with the drugs and cytokines. GFP fluorescence was read on days 1, 4 and 7 using a SpectraMax M5e Microplate Reader.

Tyrosine kinase phosphorylation profiling. Luminex immunosandwich assays were performed as previously described[29] with the following modifications. Antibodies were conjugated to MagPlex Microspheres (Luminex). The assays were carried out in 384-well Thermo Scientific Matrix square bottom plates

using a 96-well liquid handler (CyBio) and a 384-well liquid handler (Biomek). The data were acquired with a FLEXMAP 3D instrument (Luminex) according to the manufacturer's instructions. The raw data were normalized by subtracting sample and antibody backgrounds.

Immunohistochemistry. Deparaffinized tissue sections were treated with Antigen Retrieval Citra Solution (Biogenex Laboratories, HK086-9K) in a microwave for 15 min. Tissue sections were then incubated with Peroxidase Blocking Reagent (Dako, S2001) for 15 min and Protein Block (Dako, X0909) for 15 min. Primary antibody specific for HGF (R&D Systems, AB-294-NA; 0.75 $\mu g\,ml^{-1}$), MET (Invitrogen, 187366; 4 $\mu g\,ml^{-1}$), pERK (Cell Signaling Technology, 4376; 1:200 dilution) or pAKT (Cell Signaling Technology, 4060; 1:50 dilution) was applied, and the slides were incubated for 16 h at 4 °C. For HGF, the sections were then incubated with rabbit anti-goat antibody (Vector Labs, BA-5000) for 30 min. Signals were visualized using EnVision+ HRP rabbit (for HGF; Dako, K4003) or mouse (for MET; Dako, K4001) or SignalStain Boost IHC Detection Reagent (for pERK and pAKT; Cell Signaling Technology, 8114), plus diaminobenzidine (Dako, K3468) and a haematoxylin counterstain. To detect MET expression in melanoma, a VIP Peroxidase Substrate Kit (Vector Labs, SK-4600) was used instead of diaminobenzidine. All immunostained slides were scored by a pathologist (T.M.) who was blinded to the clinical outcome data.

Immunofluorescence. Fresh-frozen tissue sections were stained using the general protocol recommended by Cell Signaling Technology. Sections were blocked for 1 h in PBS containing 5% normal goat serum (Cell Signaling Technology, 5425S) and 0.3% Triton X-100. Primary antibody specific for pMET (Tyr 1234/1235) (Cell Signaling Technology, 3077S; 1:100 dilution) diluted in PBS containing 1% BSA and 0.3% Triton X-100 was applied, and slides were incubated for 16 h at 4 °C Sections were then washed with PBS and incubated with goat anti-rabbit IgG conjugated to DyLight 488 (Thermo Fisher Scientific, 35552; 1:500 dilution) for 1 h. Slides were mounted with ProLong Antifade Reagent with 4′,6-diamidino-2-phenylindole (DAPI) (Life Technologies, P36935). Experiments were carried out in parallel with SignalSlide Phospho-Met (Tyr 1234/1235) IHC Control Slides (Cell Signaling Technology, 8118) for proper staining. Images were captured using an ECLIPSE 80i fluorescence microscope (Nikon).

Calculating excess over Bliss. The Bliss independence model predicts the combined response, C, for two single compounds with response A and B, according to the relationship $C = X - [A + B - (A \times B)]$, where X is the fractional inhibition for combined compounds at given concentrations, A is the fractional inhibition of compound A at the particular concentration and B is the fractional inhibition of compound B at the particular concentration. According to this model, the excess above the predicted Bliss independence represents the synergistic effect of the combination treatment[30].

28. Chan, A. T., Ogino, S. & Fuchs, C. S. Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. *N. Engl. J. Med.* **356,** 2131–2142 (2007).
29. Du, J. *et al.* Bead-based profiling of tyrosine kinase phosphorylation identifies SRC as a potential target for glioblastoma therapy. *Nature Biotechnol.* **27,** 77–83 (2009).
30. Keith, C. T., Borisy, A. A. & Stockwell, B. R. Multicomponent therapeutics for networked systems. *Nature Rev. Drug Discov.* **4,** 71–78 (2005).

doi:10.1038/nature11249

Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors

Timothy R. Wilson[1], Jane Fridlyand[2], Yibing Yan[3], Elicia Penuel[3], Luciana Burton[3], Emily Chan[1], Jing Peng[1], Eva Lin[1], Yulei Wang[3], Jeff Sosman[4], Antoni Ribas[5], Jiang Li[6], John Moffat[7], Daniel P. Sutherlin[8], Hartmut Koeppen[9], Mark Merchant[1], Richard Neve[1] & Jeff Settleman[1]

Mutationally activated kinases define a clinically validated class of targets for cancer drug therapy[1]. However, the efficacy of kinase inhibitors in patients whose tumours harbour such alleles is invariably limited by innate or acquired drug resistance[2,3]. The identification of resistance mechanisms has revealed a recurrent theme—the engagement of survival signals redundant to those transduced by the targeted kinase[4]. Cancer cells typically express multiple receptor tyrosine kinases (RTKs) that mediate signals that converge on common critical downstream cell-survival effectors—most notably, phosphatidylinositol-3-OH kinase (PI(3)K) and mitogen-activated protein kinase (MAPK)[5]. Consequently, an increase in RTK-ligand levels, through autocrine tumour-cell production, paracrine contribution from tumour stroma[6] or systemic production, could confer resistance to inhibitors of an oncogenic kinase with a similar signalling output. Here, using a panel of kinase-'addicted' human cancer cell lines, we found that most cells can be rescued from drug sensitivity by simply exposing them to one or more RTK ligands. Among the findings with clinical implications was the observation that hepatocyte growth factor (HGF) confers resistance to the BRAF inhibitor PLX4032 (vemurafenib) in *BRAF*-mutant melanoma cells. These observations highlight the extensive redundancy of RTK-transduced signalling in cancer cells and the potentially broad role of widely expressed RTK ligands in innate and acquired resistance to drugs targeting oncogenic kinases.

Using 41 human-tumour-derived cell lines with previously defined kinase dependency[7–9], we undertook a 'matrix analysis' to examine the effects of six different RTK ligands known to be widely expressed in tumours[10] (HGF, epidermal growth factor (EGF), fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), neuregulin 1 (NRG1) and insulin-like growth factor (IGF)) on drug response. We quantified the effect of exposing these cell lines to each ligand on the half-maximum inhibitory concentration (IC_{50}) for a kinase inhibitor that otherwise potently suppresses their growth (Supplementary Fig. 1a). Nearly all of the cell lines tested, representing multiple tissue types and distinct kinase dependencies, could be rescued from drug-induced growth inhibition by one or more RTK ligands (Fig. 1a).

The consequences of ligand exposure on drug response were categorized as follows: (1) 'no rescue', addition of ligand did not affect drug response; (2) 'partial rescue', ligand partially abrogated treatment response; or (3) 'complete rescue', ligand 'right-shifted' the IC_{50} curve >10-fold, or completely suppressed drug response (Fig. 1b). HGF, FGF and NRG1 were the most broadly active ligands, followed by EGF, whereas IGF and PDGF had relatively little effect, despite activating their corresponding receptors (Supplementary Figs 1b and 5a). Many tested cell lines could be rescued from treatment sensitivity by exposure to as many as four different ligands, highlighting the capacity of such cells to engage redundant survival pathways upon exposure to a variety of ligands. None of the tested ligands could rescue

a

Kinase addiction (effective drug)	Cell line	HGF	FGF	IGF	PDGF	NRG1	EGF
HER2 amplified (lapatinib)	KPL4	P	P	NR	NR	R	NR
	EFM192A	NR	NR	NR	NR	R	NR
	SKBR3	NR	NR	NR	NR	R	R
	BT474	NR	NR	NR	NR	R	P
	AU565	P	P	NR	NR	R	P
	HCC1419	NR	NR	NR	NR	R	NR
	UACC893	NR	P	NR	NR	R	R
	HCC1954	P	NR	NR	NR	NR	P
	HCC202	NR	P	NR	NR	R	NR
NRG1 autocrine (lapatinib)	CHL-1	R	R	P	NR	NR	NR
	SAT	R	R	NR	NR	NR	R
	FaDu	R	R	NR	NR	NR	P
	PCI-6A	R	R	NR	NR	NR	R
EGFR mutation (erlotinib)	PC9	R	R	P	NR	NR	NR
	HCC827	R	P	NR	NR	NR	NR
	HCC4006	R	R	NR	NR	P	NR
	KHM-3S	R	R	P	NR	NR	NR
EGF-like ligand autocrine (erlotinib)	CALU-3	P	NR	NR	NR	NR	NR
	H358	NR	NR	NR	NR	NR	NR
	H1648	NR	NR	NR	NR	NR	NR
	H1666	R	R	NR	NR	R	NR
MET amplified (crizotinib)	KATO II	NR	R	NR	NR	R	R
	GTL16	NR	P	NR	NR	R	R
	EBC-1	NR	NR	NR	NR	P	P
PDGFR amplified (sunitinib)	NCI-H1703	R	R	NR	NR	NR	R
	A-204	NR	R	NR	NR	NR	NR
FGFR amplified (PD173074)	MFM223	NR	P	NR	NR	R	P
	Kato III	NR	NR	NR	NR	NR	NR
	RT112	R	NR	NR	NR	R	R
	RT4	NR	NR	NR	NR	R	R
ALK translocation/ mutation (TAE684)	KELLY	NR	P	NR	NR	NR	R
	SH-SY5Y	NR	P	NR	NR	NR	NR
	NCI-H3122	R	NR	NR	NR	R	R
	NCI-H2228	NR	NR	NR	NR	NR	NR
BRAF(V600E) (PLX4032)	M14	NR	P	NR	NR	P	NR
	NAE	R	NR	NR	NR	R	P
	A-375	P	R	NR	NR	P	NR
	SK-MEL-28	P	NR	NR	NR	R	NR
	Malme-3M	P	R	NR	NR	NR	P
	COLO-201	P	NR	NR	NR	NR	NR
	COLO-818	P	NR	NR	NR	NR	NR

b

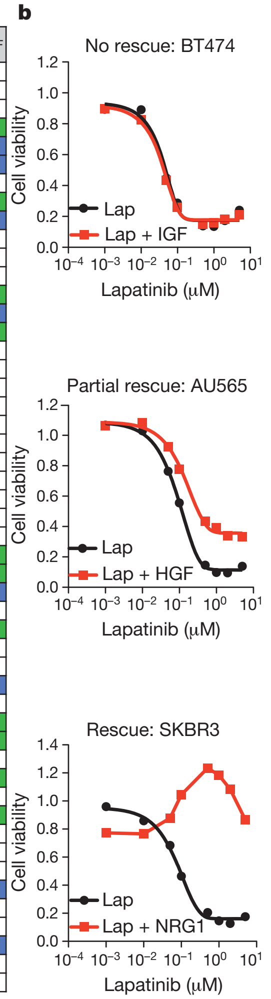


Figure 1 | RTK ligands attenuate kinase inhibition in oncogene-addicted cancer cell lines. **a**, Summary of results from 41 kinase-addicted cancer cell lines co-treated with the appropriate kinase inhibitor and each of six RTK ligands. NR, no rescue; P, partial rescue; R, complete rescue. **b**, Cell viability assay demonstrating the diversity of ligand effects on drug-treated cell lines (72 h), illustrating examples of no rescue, partial rescue or complete rescue. Lap, lapatinib. Graphs show average values of technical duplicates from one representative experiment out of three independent experiments.

[1]Research Oncology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [2]Biostatastics, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [3]Development Sciences, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [4]Vanderbilt University Medical Center, Division of Hematology/Oncology, 777 Preston Research Building, Nashville, Tennessee 37232, USA. [5]Jonsson Comprehensive Cancer Center, University of California Los Angeles, 9-954 Factor Building, 10833 Le Conte Avenue, Los Angeles, California 90095, USA. [6]Medical Developmental Biometrics Biostatistics, Roche Nutley, 340 Kingsland Street, Nutley, New Jersey 07110, USA. [7]Biochemical and Cellular Pharmacology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [8]Discovery Chemistry, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [9]Research Pathology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA.

cells from cisplatin sensitivity, suggesting that RTK ligands do not confer broad protection from toxic agents (Supplementary Fig. 1c).

To explore the signalling dynamics associated with ligand-mediated rescue, we assessed the status of two downstream survival signalling pathways commonly engaged by RTKs: the PI(3)K–AKT and MAPK pathways[11]. Where ligand-mediated rescue was achieved, ligand re-activated at least one of these pathways despite the presence of drug (Fig. 2a). Pathway re-activation was not due to re-activation of the oncogenic kinase, as kinase autophosphorylation remained suppressed after ligand co-treatment. In the tested models, HGF re-activated both PI(3)K and MAPK, IGF and NRG1 only re-activated PI(3)K and FGF, and EGF only re-activated MAPK. Activation of the 'redundant RTK' and consequent downstream survival signalling persisted for at least 48 h (Supplementary Fig. 1d). A functional role for re-activation of both PI(3)K and MAPK signalling was observed in lapatinib-treated AU565 cells in the presence of NRG1, FGF or a combination (Supplementary Fig. 2a). However, specifically inhibiting PI(3)K attenuated HGF-promoted drug resistance, which was associated with engagement of both survival pathways (Supplementary Fig. 2b). As expected, ligand rescue was blocked by co-targeting the secondary activated kinase, confirming that effective ligands were acting via their cognate RTKs (Fig. 2b, c and Supplementary Fig. 3a, b). Inhibitors of the 'secondary' RTK that mediated ligand rescue had little or no effect as single agents, indicating that the kinase-addicted cells are not initially dependent on multiple different RTKs in the absence of available ligand. Moreover, ligand stimulation had little or no effect on cell proliferation (Figs 1b and 2b).

Analysis of baseline RTK expression confirmed that the tested lines express multiple RTKs. Ligand-induced rescue was well correlated with expression of certain RTKs in some cases (Supplementary Fig. 4), suggesting that the RTK profile of tumours before treatment could inform an optimal treatment strategy that anticipates the need to co-target two or more kinases. In some cases, ligands were unable to rescue cells from drug effects despite expression of the corresponding RTK. In a few cases, the RTK ligand activated its receptor; however, engagement of PI(3)K or MAPK was not observed (Supplementary Fig. 5a). In other cases, the ligand activated its receptor as well as at least one downstream effector; however, that was not sufficient for rescue (Supplementary Fig. 5b, c). This was observed, for example, with H2228 and H358 cells exposed to HGF, or with COLO-201 cells exposed to NRG1. However, H2228 and H358 cells are rescued by HGF after longer-term drug treatment, implicating the presence of a subpopulation of HGF-responsive cells selected over time in the presence of an inhibitory kinase (Supplementary Fig. 6c, d).

There were several findings with clinical implications. For example, one of two tested non-small-cell lung carcinoma (NSCLC) cell lines harbouring an anaplastic lymphoma kinase (ALK) translocation (NCI-H3122 cell line), and exhibiting ALK addiction, was rescued

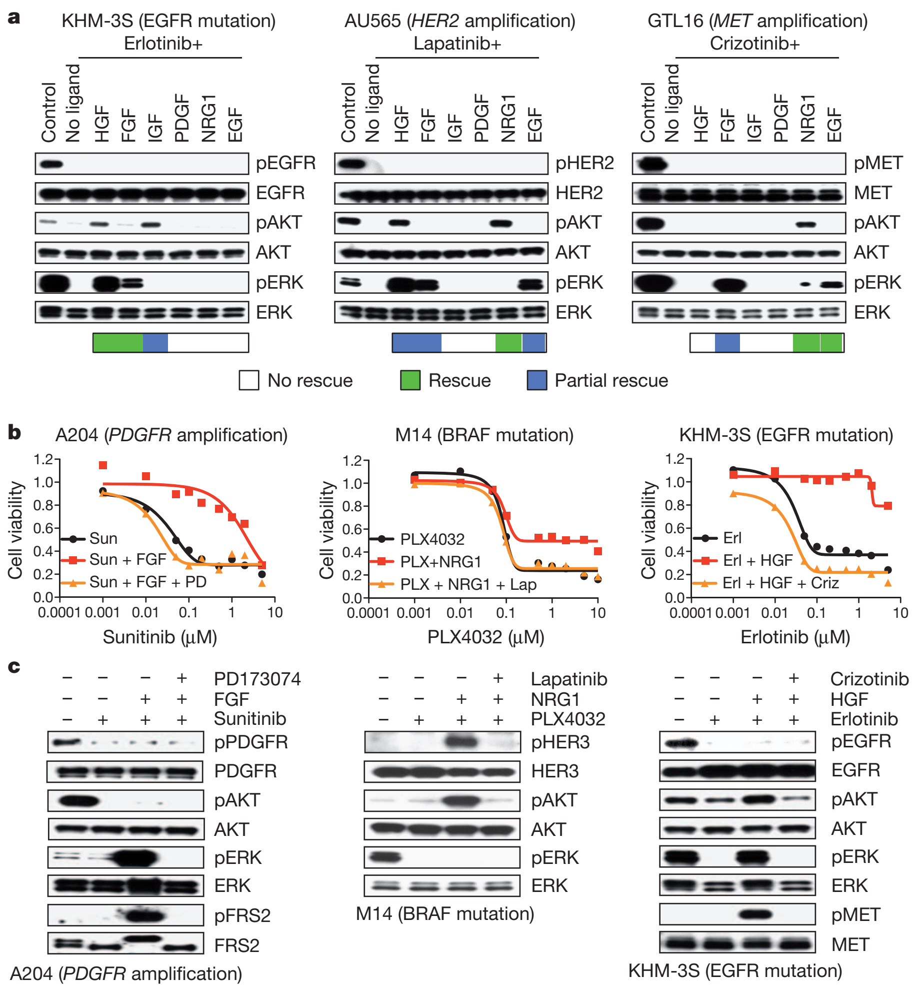


Figure 2 | Pro-survival pathway re-activation correlates with RTK-ligand rescue. **a**, Immunoblots showing effects of ligand (50 ng ml^{-1}) on RTK, AKT and extracellular signal-regulated kinase (ERK) phosphorylation (p) after kinase inhibition (1 μM, 2 h). Ligand rescue is indicated: green squares, complete rescue; blue squares, partial rescue. **b**, Drug-induced suppression of viability in three kinase-addicted cell lines (72 h). Cells were co-treated with ligand and an appropriate secondary kinase inhibitor (0.5 μM). Graphs show average values of technical duplicates from one representative experiment out of three independent experiments. Criz; crizotinib; Erl, erlotinib; Lap, lapatinib; PD, PD173074; PLX, PLX4032; Sun, sunitinib. **c**, Immunoblots showing effect of kinase inhibition (1 μM) ± ligands (2 h) on RTK, AKT and ERK phosphorylation. Cells were co-treated with secondary kinase inhibitor (0.5 μM).

from ALK inhibition by brief exposure to HGF (Supplementary Fig. 6a, b). In these HGF receptor (MET)-expressing cells, HGF activates extracellular signal-regulated kinase (ERK) and AKT even in the presence of the ALK-selective inhibitor TAE684. In contrast, survival of these cells was disrupted even in the presence of HGF by crizotinib, a dual ALK/MET kinase inhibitor approved for treatment of ALK-translocated NSCLCs[12]. The second ALK-translocated NSCLC line, NCI-H2228, also expresses MET, but was not rescued from ALK inhibition by HGF at the 72-h time point. However, HGF re-activated AKT and ERK in the presence of TAE684 (Supplementary Fig. 5b), and longer-term TAE684 treatment in the presence of HGF yielded TAE684-resistant cells (Supplementary Fig. 6c). The relatively durable clinical responses observed in ALK-translocated NSCLC patients might be attributed in part to the dual ALK/MET inhibitory activity of crizotinib.

The ability of HGF to rescue three of nine tested human epidermal growth factor receptor 2 (*HER2*)-amplified breast cancer cell lines from growth inhibition by the HER2 kinase inhibitor lapatinib was also unexpected (Fig. 3a, b). These three cell lines express MET, which was correlated with the ability of HGF to attenuate lapatinib response (Fig. 3b). As with NCI-H228 cells, longer-term co-treatment (12 days) of the partially HGF-rescued AU565 cells revealed that HGF rapidly promoted lapatinib resistance, potentially by driving selection of a subpopulation of MET-expressing cells (Fig. 3c). Indeed, 9-day lapatinib/HGF co-treatment yielded a population of cells with increased MET (Supplementary Fig. 7a). HGF re-activated PI(3)K and MAPK signalling specifically in MET-positive cells (Fig. 3d).

A subset of tested HER2-positive primary breast tumours express MET protein (Supplementary Fig. 7b). One *HER2*-amplified breast cancer cell line (HCC1954) showed elevated phospho-MET in the absence of exogenous HGF, implicating an autocrine mechanism (Fig. 3b), and MET inhibition in these cells delayed the emergence of lapatinib resistance (Fig. 3e). Collectively, these results suggest that MET-expressing HER2-positive breast tumours might evade HER2 inhibition by engaging MET in a subpopulation of 'primed' tumour cells, depending on the availability of HGF. Additionally, eight of nine tested *HER2*-amplified breast cell lines were rescued from lapatinib sensitivity by exposure to the HER3 ligand NRG1 (Fig. 1a and Supplementary Fig. 7c), possibly implicating NRG1 levels in the tumour microenvironment in the variable response to HER2 kinase inhibition.

Another observation with clinical implications was the unexpected finding that HGF attenuated the response to the BRAF kinase inhibitor PLX4032 (vemurafenib) in *BRAF*-mutant cell lines. PLX4032 recently demonstrated remarkable efficacy in *BRAF*-mutant melanoma, leading to clinical approval[13]. Notably, among 446 tested secreted factors, HGF was among a very small number that could rescue *BRAF*-mutant melanoma cells from PLX4032 sensitivity (Supplementary Fig. 8).

Among 12 additional *BRAF*-mutant melanoma cell lines tested, HGF significantly attenuated PLX4032 sensitivity in 5 lines (Fig. 4a). Eight of ten HGF-rescued cell lines expressed detectable MET, whereas MET was undetectable or barely detectable in the non-rescued cells. MET expression was correlated with HGF rescue (Fig. 4a), and HGF re-activated MAPK in lines rescued by HGF, but not in MET-negative HGF-non-rescued cells (Fig. 4b). As anticipated, rescue by HGF was blocked when MET was inhibited by crizotinib (Fig. 4b and Supplementary Fig. 9a). One *BRAF*-mutant cell line (624MEL) showed elevated phospho-MET in the absence of exogenous HGF, consistent with an autocrine mechanism (Fig. 4a), and MET inhibition in these cells delayed PLX4032 resistance (Supplementary Fig. 9b). Crizotinib co-treatment also prevented resistance to PLX4032 in two cell lines (A375 and 928MEL) with undetectable phospho-MET, further supporting a role for HGF-activated MET in PLX4032 resistance (Supplementary Fig. 9b).

To confirm these findings *in vivo*, we performed xenograft studies with *BRAF*-mutant 928MEL and 624MEL melanoma cells. Activation of MET in tumours using the MET-agonistic antibody 3D6 strongly abrogated the growth-suppressive effects of PLX4032 (Fig. 4c). The relevance of MET activation by 3D6 in attenuating the PLX4032 response was verified by co-treating with a MET kinase inhibitor. Inhibiting MET enhanced the effect of PLX4032 on tumour regression (Fig. 4c), with more partial responses observed in the 928MEL model (Supplementary Fig. 10).

To extend these findings in a clinical context, we tested the hypothesis that circulating HGF in *BRAF*-mutant melanoma patients could contribute to clinical outcome. Pre-treatment plasma HGF levels were

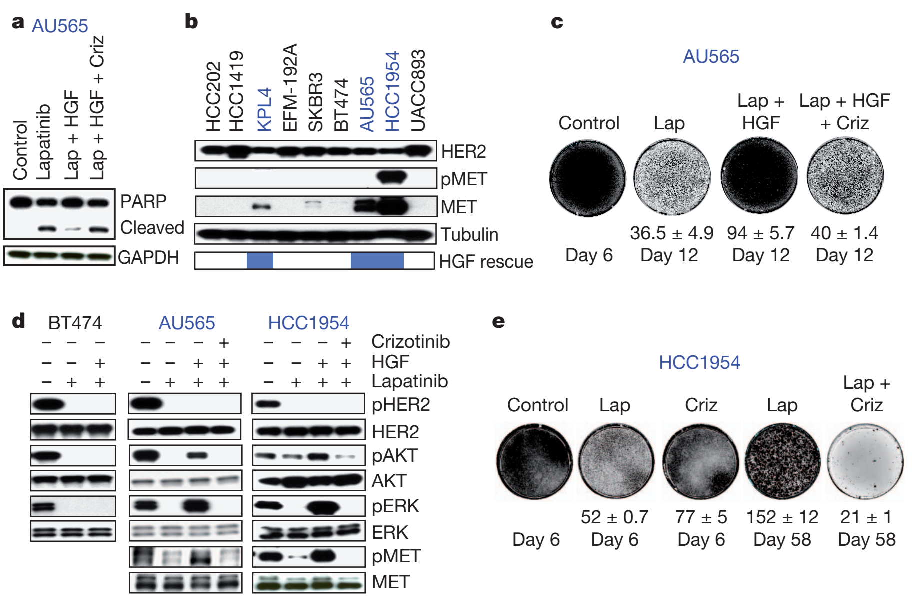


Figure 3 | HGF promotes lapatinib resistance in *HER2*-amplified breast cancer cell lines. **a**, Immunoblots showing apoptosis in AU565 cells after lapatanib (Lap; 1 μM), HGF or crizotinib (Criz; 0.5 μM) treatment. Blue font indicates partially rescued cell lines. **b**, Immunoblots showing phospho-MET (pMET) and MET in *HER2*-amplified breast cancer cell lines with partial HGF rescue indicated. **c**, Syto 60 cell staining of AU565 cells treated with lapatinib (1 μM), HGF or crizotinib (0.5 μM). **d**, Immunoblots showing AKT and ERK re-activation in MET-positive (blue) and MET-negative (black) cells. Cells were treated with lapatinib (1 μM), HGF or crizotinib (0.5 μM) (2 h). **e**, Syto 60 staining of HCC1954 cells treated with lapatinib (5 μM) or crizotinib (1 μM). Images are representative of three biological replicates and values indicate mean ± s.d.

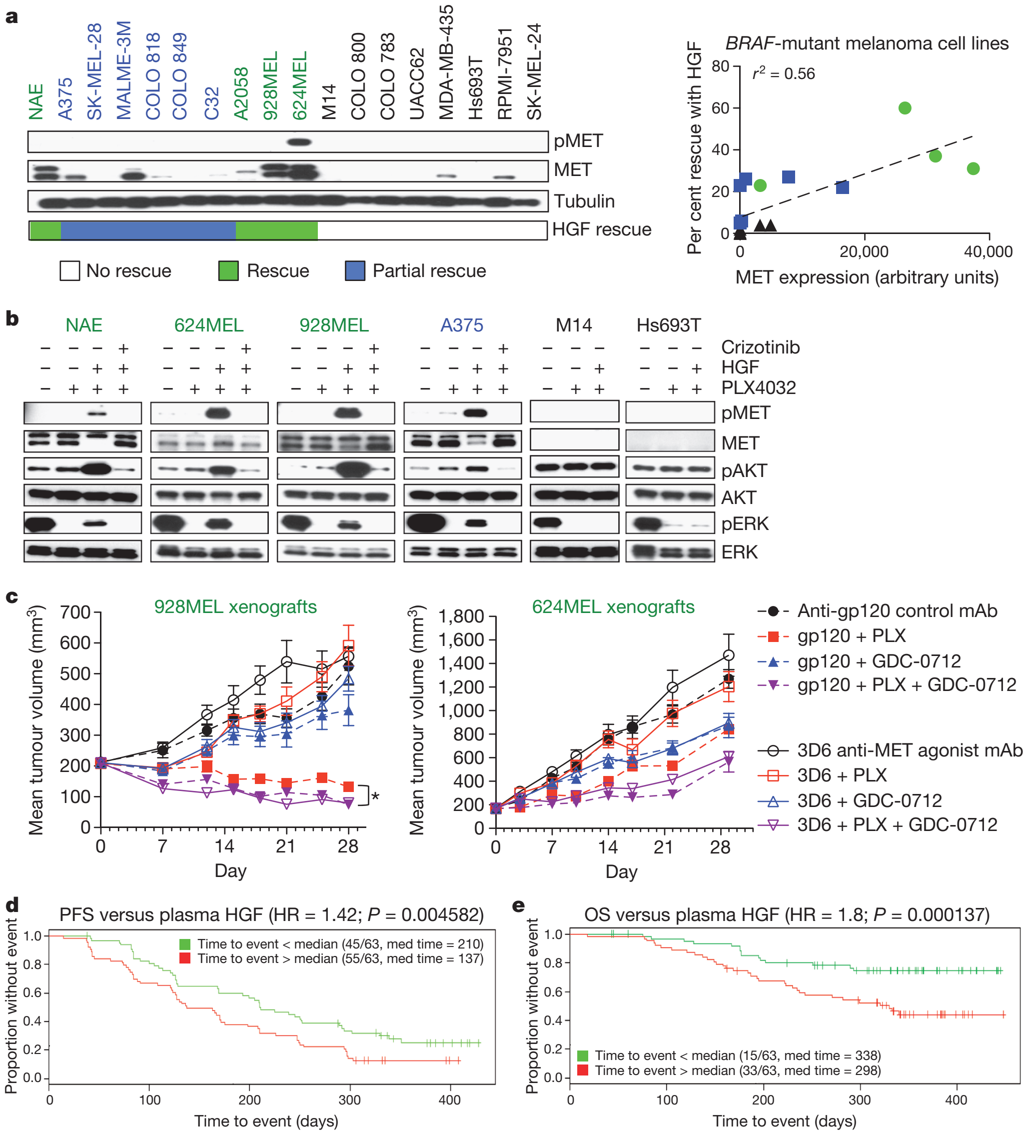


Figure 4 | HGF promotes PLX4032 resistance in *BRAF*-mutant melanoma cell lines. **a**, Left, immunoblots showing MET status in melanoma cells, with HGF rescue indicated. Right, correlation between MET expression and HGF rescue in PLX4032 (1 μM)-treated cells (72 h). Black font indicates no rescue; blue, partial rescue; green, complete rescue. **b**, Immunoblots showing ERK re-activation in MET-positive (green/blue) and MET-negative (black) cells treated with PLX4032 (1 μM), HGF or crizotinib (0.5 μM) (2 h). **c**, Effect of activating MET (3D6) on tumour growth inhibition by PLX4032 in two xenografts ($n = 10$ per group). mAb, monoclonal antibody; PLX, PLX4032. Differences between PLX4032-treated and PLX4032- and GDC-0712 (MET inhibitor)-treated control antibody (gp120) groups (*$P = 0.0008$). Error bars represent mean ± s.e.m. (biological replicates). **d**, **e**, PFS (**d**) and OS (**e**) in PLX4032-treated melanoma patients stratified based on plasma HGF (green < median HGF; red > median HGF).

measured from 126 of the 132 patients enrolled onto the BRIM2 clinical trial (*BRAF*-mutant metastatic melanoma patients treated with PLX4032). HGF levels ranged from 33–7,200 pg ml^{-1} with a median level of 334 pg ml^{-1} (Supplementary Fig. 11a). Increased plasma HGF was associated with worse outcome as measured by progression-free survival (PFS; hazard ratio, 1.42; $P < 0.005$) and overall survival (OS; hazard ratio, 1.8; $P < 0.001$; Fig. 4d, e). Segregating patients into tertiles revealed a continuous relationship between HGF level and outcome, rather than a threshold effect (Supplementary Fig. 11b). As BRIM2 was a single-arm study in which all patients received PLX4032, it is not possible to determine whether higher HGF levels confer drug resistance; however, this study implicates HGF–MET signalling in disease progression and overall survival, and together with related findings[14], suggests a potential role for HGF in the response to BRAF inhibition in *BRAF*-mutant melanoma.

Overall, the findings highlight the extensive nature of signal cross-talk among RTKs that are co-expressed in most tumour cells, and the potentially broad role of RTK ligands in innate and acquired resistance to kinase inhibitors. Such ligands could be produced by tumour cells themselves to drive autocrine survival mechanisms, by tumour stroma, or systemically to affect drug response[15,16].

The increasingly appreciated heterogeneity of human tumours complicates the elucidation of drug resistance mechanisms[17–19]. In the context of our findings, we imagine distinct mechanisms by which such heterogeneity could contribute to acquired resistance. Thus, a subpopulation of tumour cells present before therapy capable of responding to a survival-promoting RTK ligand might be expanded through the selective pressure of drug treatment if such a ligand becomes available within the tumour microenvironment. Indeed, analysis of MET expression in *BRAF*-mutant melanoma cells revealed a heterogeneous cell population (Supplementary Fig. 12). Similarly, in *EGFR*-mutant NSCLC, a subpopulation of MET-driven tumour cells may emerge upon exposure to HGF during EGFR tyrosine kinase inhibitor treatment[20]. Activation of multiple RTKs has been reported in glioblastoma, and suppression of pro-survival signals and consequent cell death was only observed after co-targeting multiple

activated RTKs[21]. A subpopulation of RTK-ligand-producing tumour cells may be selected during treatment. In various pre-clinical models, the observed acquired resistance mechanism involved a 'switch' to a new RTK dependency[22–27], which in some cases could be attributed to increased RTK-ligand production.

Although genomic biomarkers have been critical in identifying the patients most likely to benefit from therapy, there is a largely unexplained wide range of clinical benefit among such patients in terms of magnitude and duration of response[12,13]. Our findings support a potentially broad role for RTK ligands in the overall clinical benefit from such therapies, and provide a foundation for the use of biomarkers based on the expression of RTKs and their ligands to inform treatment strategies that anticipate both innate and acquired resistance mechanisms.

METHODS SUMMARY

RTK-ligand matrix screen. Cell viability was assessed using the nucleic acid stain Syto 60 (Invitrogen). Cells (3,000–5,000 per well) were seeded into 96-well plates. The next day, cells were treated with (or without) $50\,\mathrm{ng\,ml^{-1}}$ RTK ligand and concomitantly exposed to increasing concentrations of the relevant kinase inhibitor. After 72 h, cells were fixed in 4% formaldehyde, stained with Syto 60 and cell number was assessed using an Odyssey scanner (Li-Cor). Cell viability was calculated by dividing the fluorescence from the drug-treated cells by the fluorescence from the control (no drug) treated cells.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 November 2011; accepted 23 May 2012.
Published online 4 July 2012.

1. Sharma, S. V. & Settleman, J. Oncogene addiction: setting the stage for molecularly targeted cancer therapy. *Genes Dev.* **21,** 3214–3231 (2007).
2. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3,** 75ra26 (2011).
3. Garrett, J. T. & Arteaga, C. L. Resistance to HER2-directed antibodies and tyrosine kinase inhibitors: mechanisms and clinical implications. *Cancer Biol. Ther.* **11,** 793–800 (2011).
4. Engelman, J. A. & Settleman, J. Acquired resistance to tyrosine kinase inhibitors during cancer therapy. *Curr. Opin. Genet. Dev.* **18,** 73–79 (2008).
5. Moritz, A. *et al.* Akt–RSK–S6 kinase signaling networks activated by oncogenic receptor tyrosine kinases. *Sci. Signal.* **3,** ra64 (2010).
6. Zhang, W. & Huang, P. Cancer-stromal interactions: role in cell survival, metabolism and drug sensitivity. *Cancer Biol. Ther.* **11,** 150–156 (2011).
7. McDermott, U. *et al.* Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc. Natl Acad. Sci. USA* **104,** 19936–19941 (2007).
8. Wilson, T. R., Lee, D. Y., Berry, L., Shames, D. S. & Settleman, J. Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. *Cancer Cell* **20,** 158–172 (2011).
9. Yonesaka, K. *et al.* Autocrine production of amphiregulin predicts sensitivity to both gefitinib and cetuximab in EGFR wild-type cancers. *Clin. Cancer Res.* **14,** 6963–6973 (2008).
10. Mueller, M. M. & Fusenig, N. E. Friends or foes — bipolar effects of the tumour stroma in cancer. *Nature Rev. Cancer* **4,** 839–849 (2004).
11. Grant, S., Qiao, L. & Dent, P. Roles of ERBB family receptor tyrosine kinases, and downstream signaling pathways, in the control of cell growth and survival. *Front. Biosci.* **7,** d376–d389 (2002).
12. Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363,** 1693–1703 (2010).
13. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364,** 2507–2516 (2011).
14. Straussman, R. *et al.* Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* http://dx.doi.org/10.1038/nature11183 (2012).
15. Gilbert, L. A. & Hemann, M. T. DNA damage-mediated induction of a chemoresistant niche. *Cell* **143,** 355–366 (2010).
16. Liang, K. *et al.* Recombinant human erythropoietin antagonizes trastuzumab treatment of breast cancer cells via Jak2-mediated Src activation and PTEN inactivation. *Cancer Cell* **18,** 423–435 (2010).
17. Quintana, E. *et al.* Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. *Cancer Cell* **18,** 510–523 (2010).
18. Calbo, J. *et al.* A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* **19,** 244–256 (2011).
19. Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141,** 69–80 (2010).
20. Turke, A. B. *et al.* Preexistence and clonal selection of *MET* amplification in *EGFR* mutant NSCLC. *Cancer Cell* **17,** 77–88 (2010).
21. Stommel, J. M. *et al.* Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318,** 287–290 (2007).
22. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468,** 973–977 (2010).
23. Guix, M. *et al.* Acquired resistance to EGFR tyrosine kinase inhibitors in cancer cells is mediated by loss of IGF-binding proteins. *J. Clin. Invest.* **118,** 2609–2619 (2008).
24. McDermott, U., Pusapati, R. V., Christensen, J. G., Gray, N. S. & Settleman, J. Acquired resistance of non-small cell lung cancer cells to MET kinase inhibition is mediated by a switch to epidermal growth factor receptor dependency. *Cancer Res.* **70,** 1625–1634 (2010).
25. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468,** 968–972 (2010).
26. Liu, L. *et al.* Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. *Cancer Res.* **69,** 6871–6878 (2009).
27. Engelman, J. A. *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316,** 1039–1043 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Settleman laboratory, G. Bray and C. Bowdoin for helpful discussions, K. Trunzer and B. Nelson for assistance with access to clinical samples.

Author Contributions T.R.W. and J.S. designed the study, analysed data, discussed results and co-wrote the paper. T.R.W. performed *in vitro* experiments. T.R.W., E.L. and R.N. designed and performed the 446 soluble factor screen. H.K. performed immunohistochemistry analysis. E.C., J.P. and M.M. designed and performed *in vivo* experiments. E.P., L.B., Y.W. and Y.Y. assessed BRIM2 study material, including HGF enzyme-linked immunosorbent assay (ELISA) from plasma. J.F. carried out the biostatistical analysis. J.L. carried out efficacy and safety analyses on the BRIM2 study. J.S. and A.R. were clinical investigators on the BRIM2 study. J.M. and D.P.S. characterized GDC-0712.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S. (settleman.jeffrey@gene.com).

METHODS

Cell lines. Human cancer cell lines were obtained and tested for sensitivity using an automated platform as previously described[7]. Cell lines were maintained at 37 °C in a humidified atmosphere at 5% CO_2 and grown in RPMI 1640 or DMEM/F12 growth media (GIBCO) supplemented with 10% fetal bovine serum (GIBCO), 50 units ml^{-1} penicillin and 50 μg ml^{-1} streptomycin (GIBCO).

Reagents. Lapatinib, sunitinib and erlotinib were purchased from LC Laboratories. Crizotinib, TAE684, AZD6244 and BEZ235 were purchased from Selleck Chemicals. PD173074 was purchased from Tocris Bioscience. PLX4032 was purchased from Active Biochem. Recombinant human HGF, EGF, FGF-basic, IGF-1 and PDGF-AB were purchased from Peprotech. Recombinant human NRG1-β1 was purchased from R&D Systems. For *in vivo* studies, 3D6 anti-MET agonist antibody, PLX4032 and GDC-0712 were generated at Genentech. GDC-0712 was used in xenograft experiments as it has a similar kinase profile as crizotinib[28] (Supplementary Fig. 13) and was available in quantities sufficient for *in vivo* study. See Supplementary Methods for synthesis protocol.

Immunoblot analysis. Cell lysates were collected using Nonidet-P40 lysis buffer, supplemented with Halt protease and phosphatase inhibitor cocktail (Thermo Scientific) and immunodetection of proteins was carried out using standard protocols. The phospho-HER2 (Y1248; catalogue no. 2247), HER2 (no. 2242), phospho-HER3 (Y1289; no. 4791), phospho-MET (Y1234/5; no. 3126), PDGFRα (no. 5241), phospho-FRS2α (Y196; no. 3864), IGF-1Rβ (no. 3027), phospho-ALK (Y1604; no. 3341), AKT (no. 9272), phospho-ERK (T202/Y204; no. 9101), ERK (no. 9102), GAPDH (no. 2118) and β-tubulin (no. 2146) antibodies were purchased from Cell Signaling Technologies. Antibodies to HER3 (SC-285), MET (SC-10), phospho-PDGFRα (SC-12911), FRS2α (SC-8318), FGFR1 (SC-7945), FGFR2 (SC-122), FGFR3 (SC-13121) and ALK (SC-25447) were purchased from Santa Cruz Biotechnologies. Phospho-AKT (S473; no. 44-621G) antibody was purchased from Invitrogen. Phospho-EGFR (Y1068; ab5644) antibody was purchased from Abcam. EGFR (no. 610017) antibody was purchased from BD Biosciences. PARP (no. 14-6666-92) antibody was purchased from eBioscience. Densitometry was carried out using ImageJ software.

Tissue samples. Primary breast tumour samples with appropriate Institutional Review Board (IRB) approval and informed patient consent were obtained from the following sources: Cureline, ILSbio and the Cooperative Human Tissue Network of the National Cancer Institute. Metastatic melanoma tumour samples with appropriate IRB approval and informed patient consent were obtained from the BRIM2 trial. The human tissue samples used in the study were de-identified (double-coded) before their use and thus the study using these samples is not considered human subject research under the US Department of Human and Health Services regulations and related guidance (45 CFR, Part 46). Immunohistochemistry for MET was performed on formalin-fixed paraffin-embedded sections cut at a thickness of 4 μm on to positively charged glass slides. The staining was performed on a Discovery XT autostainer with Ultraview detection (VMSI) using the MET rabbit monoclonal antibody SP44 (Spring BioScience; no. M3441) and CC1 standard antigen retrieval. Sections were counterstained with haematoxylin and specific membranous staining for MET was scored on a scale from 0 (no staining) to 3+ (strong staining).

HGF ELISA. Plasma was obtained from a metastatic melanoma patient's pre-dose PLX4032 cycle one and the concentration of HGF in patient-derived plasma was quantitatively measured using a sandwich ELISA as previously described[29].

Xenograft studies. All procedures were approved by and conformed to the guidelines and principles set by the Institutional Animal Care and Use Committee of Genentech and were carried out in an Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC)-accredited facility. Ten million 928MEL or 624MEL *BRAF*-mutant melanoma cells (suspended in a 1:1 mixture of HBSS/Matrigel) were inoculated in the right flank of CRL C.B-17 SCID.bg mice (Charles River Laboratories). When tumours reached an average volume of 200 mm^3, mice (10 per group) were treated with either control antibody (anti-gp120; 10 mg kg^{-1} once per week; intraperitoneal), 3D6 (anti-MET agonistic antibody; 10 mg kg^{-1} once per week; intraperitoneal), PLX4032 (50 mg kg^{-1} twice daily; periocular), GDC-0712 (MET small molecular inhibitor[28], previously named GNE-A; 100 mg kg^{-1} every day; periocular) as indicated for 4 weeks. Tumours were measured twice weekly using digital calipers (Fred V. Fowler Company) and tumour volumes were calculated using the formula $(L \times (W \times W))/2$. A partial response was defined as a reduction in tumour volume greater than 50% but less than 100%. A complete response was defined as 100% reduction in tumour volume. Differences between the PLX4032-treated and the PLX4032- and GDC-0712-treated control antibody groups were determined using two-way ANOVA ($P = 0.0008$).

Secreted factor screen. Recombinant purified secreted factors were purchased from Peprotech and R&D Systems as indicated, and were reconstituted in PBS/0.1% BSA (Supplementary Table 1). Secreted factors were transferred into 96-well plates at a concentration of 1 μg ml^{-1}, and subsequently diluted to 100 ng ml^{-1} in media containing either no drug or 5 μM PLX4032. Equal volumes of diluted factor (final concentration 50 ng ml^{-1}) were arrayed into the 384 well plates pre-seeded with SK-MEL-28 cells (500 cells per wells seeded the day before) using an Oasis liquid handler. After 72 h incubation, cell viability was determined using Cell Titer Glo (Promega).

Statistics. Cell viability assays were carried out in duplicate wells within an individual experiment (technical replicates) and carried out multiple times (biological replicates) as indicated. Correlation of receptor expression with cognate ligand rescue was carried out using a 2×2 contingency table with the following groups: receptor positive, RTK-ligand rescued; receptor positive, RTK-ligand non-rescued; receptor negative, RTK-ligand rescued; receptor negative, RTK-ligand non-rescued. Significance was determined using a two-tailed Fisher exact probability test.

Statistical analysis of BRIM2 clinical samples. HGF levels were log transformed, and the Kolmogorov–Smirnoff test was used to test the resulting distribution for departure from the Gaussian distribution. The Cox-proportional model was used to test the log-transformed HGF levels for association with the PFS and OS. Association between the response and HGF levels was tested using the Wilcoxon rank-sum test. Kaplan–Meier curves were used to show the relationship between the HGF levels and the time-to-event outcomes (PFS and OS). The number of events/patients and medium time to event is shown for each group. The Cox-proportional model of the outcome as the function of the continuous HGF level was used to calculate the hazard ratio and corresponding *P*-value.

28. Liederer, B. M. *et al.* Preclinical absorption, distribution, metabolism, excretion, and pharmacokinetic-pharmacodynamic modelling of *N*-(4-(3-((3S,4R)-1-ethyl-3-fluoropiperidine-4-ylamino)-1*H*-pyrazolo[3,4-*b*]p yridin-4-yloxy)-3-fluorophenyl)-2-(4-fluorophenyl)-3-oxo-2,3-dihydropyrida zine-4-carboxamide, a novel MET kinase inhibitor. *Xenobiotica* **41,** 327–339 (2011).
29. Catenacci, D. V. T. *et al.* Durable complete response of metastatic gastric cancer with anti-met therapy followed by resistance at recurrence. *Cancer Discovery* **1,** 573–579 (2011).

LETTER

doi:10.1038/nature11182

Structural insights into electron transfer in caa_3-type cytochrome oxidase

Joseph A. Lyons[1,2], David Aragão[2,3], Orla Slattery[1], Andrei V. Pisliakov[4], Tewfik Soulimane[1,5] & Martin Caffrey[2,6]

Cytochrome *c* oxidase is a member of the haem copper oxidase superfamily (HCO)[1]. HCOs function as the terminal enzymes in the respiratory chain of mitochondria and aerobic prokaryotes, coupling molecular oxygen reduction to transmembrane proton pumping. Integral to the enzyme's function is the transfer of electrons from cytochrome *c* to the oxidase via a transient association of the two proteins. Electron entry and exit are proposed to occur from the same site on cytochrome *c*[2–4]. Here we report the crystal structure of the caa_3-type cytochrome oxidase from *Thermus thermophilus*, which has a covalently tethered cytochrome *c* domain. Crystals were grown in a bicontinuous mesophase using a synthetic short-chain monoacylglycerol as the hosting lipid. From the electron density map, at 2.36 Å resolution, a novel integral membrane subunit and a native glycoglycerophospholipid embedded in the complex were identified. Contrary to previous electron transfer mechanisms observed for soluble cytochrome *c*, the structure reveals the architecture of the electron transfer complex for the fused cupredoxin/cytochrome *c* domain, which implicates different sites on cytochrome *c* for electron entry and exit. Support for an alternative to the classical proton gate characteristic of this HCO class is presented.

HCO members are classified on the basis of sequence signatures and conservation of the D- and K-proton pathways, and fall into three main families[5,6]: types A, B and C. Type A HCOs can be further divided into two subclasses, according to the amino acid motif at the hydrophobic end of the D-pathway (Supplementary Fig. 1). Type A1 contains a highly conserved glutamate residue in the sequence motif –XGHP**E**V–. In type A2, the gating glutamate has been proposed to be replaced spatially by a consecutive tyrosine and serine in a –**YS**HPXV– motif[5].

The crystal structure of caa_3-type cytochrome *c* oxidase from *Thermus thermophilus* is the first for a type A2 HCO (Fig. 1). The structure was determined to 2.36 Å resolution with crystals (Supplementary Fig. 2) grown by the *in meso* method using a synthetic, short chain (14 carbon) monoacylglycerol (7.7 MAG) as the hosting lipid[7]. 7.7 MAG was chosen for the larger water channel and reduced interfacial curvature of the cubic mesophase it forms to more suitably accommodate the enzyme's extensive intra- and extramembrane domains[7,8]. Increasingly, 7.7 and related short chain MAGs are proving to be useful for *in meso* crystallization[8–11]. Initial phases were determined by molecular replacement with the mitochondrially encoded SU I, II and III, from bovine heart cytochrome oxidase (Protein Data Bank ID 2OCC), identifying two molecules (molecules 1 and 2) in the asymmetric unit. Consistent with the hypothesis for crystallization *in meso*[12], packing within the crystal was layered or type I (Supplementary Fig. 3). The cytochrome *c* domain was built manually after the haem iron was assigned by an anomalous map. The residual R_{work} and R_{free} of the refined structure were 17.1% and 21.8%, respectively (Supplementary Table 1, Methods; Supplementary Fig. 4 shows an experimental map of a portion of the oxidase structure) and the model ranked in the 99th percentile in MolProbity geometry analysis[13].

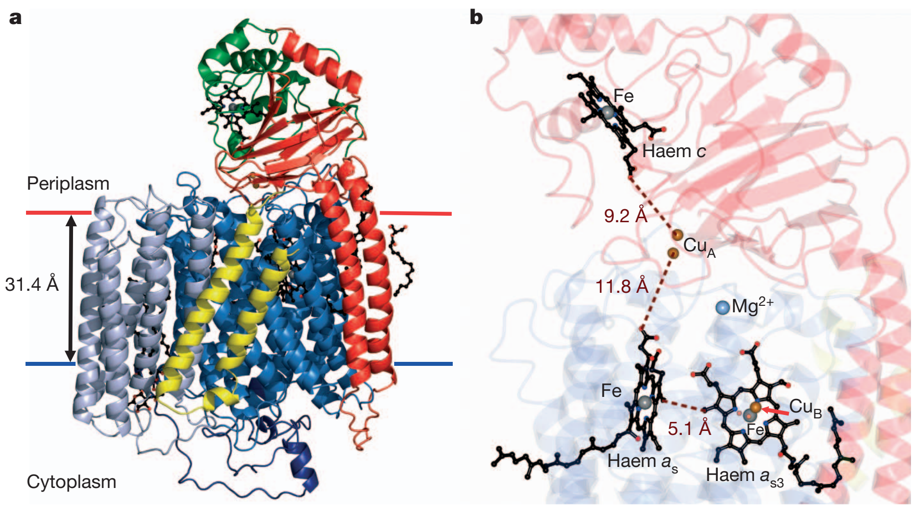


Figure 1 | Structure of and cofactor arrangement in caa_3-oxidase. **a**, Structure (ribbon model). SU I/III is coloured to highlight its canonical subunits: SU I, blue; SU III, blue-grey and the fusion linker, dark blue. SU IIc is coloured to highlight the classical SU II (red) and the fused cytochrome *c* domain (green). SU IV is in yellow. The haems are in ball and stick with the iron and copper metal centres as grey and copper spheres, respectively. Membrane boundaries are based on hydrophobic thickness calculations from the OPM server[31]. **b**, Cofactor arrangement. Haems *c*, a_s and a_{s3} and the iron and copper ions are shown as in **a**. The magnesium ion is represented as a light blue sphere. Distances are shown in brown. SU I/III, IIc and IV are colour-coded as faded blue, red and yellow ribbons, respectively.

[1]Department of Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland. [2]School of Biochemistry and Immunology, Trinity College Dublin, Dublin, Ireland. [3]Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia. [4]Theoretical Biochemistry Laboratory, RIKEN Advanced Science Institute, Wako, Saitama 351-0198, Japan. [5]Materials and Surface Science Institute, University of Limerick, Limerick, Ireland. [6]School of Medicine, Trinity College Dublin, Dublin, Ireland.

The caa_3-oxidase from *T. thermophilus* has been described as consisting of two subunits, SU I/III and IIc, that contain the metal centres characteristic of cytochrome *c* oxidases[14,15]. SU I/III exists as a fusion of the classical SU I and SU III[16], whereas SU IIc is a fusion of a canonical SU II and a cytochrome *c* domain[17]. During the course of caa_3-oxidase structure determination, electron density for a third subunit, SU IV, consisting of two connected transmembrane helices located along the hydrophobic surface of SU I/III (Fig. 1a and Supplementary Fig. 5) and a native glycoglycerophospholipid (see Supplementary Discussion and Supplementary Figs 6 and 7) embedded in SU I/III were identified. SU IV was modelled based on a 66 residue, amino-terminally formylated protein sequence determined using vapour-diffusion-grown crystals (see Methods and Supplementary Fig. 5).

The final model consists of three subunits, SU I/III, IIc and IV, which collectively form a bundle of 23 transmembrane helices and a bulky extramembrane cytochrome *c*/cupredoxin domain (Fig. 1a). SU I/III is comprised of 19 transmembrane helices, which hosts the low-spin haem, a high-spin haem–Cu_B binuclear centre (Fig. 2a, b) and a conserved magnesium binding site (Fig. 2c). The fusion linker between the SU I and III domains is approximately 70 residues long, largely devoid of secondary structure save for a two-turn helix, and drapes the cytoplasmic surface of SU I/III, possibly conferring stability to the helical bundle. SU IIc consists of a membrane anchored cupredoxin domain, containing the electron-accepting homo-binuclear copper (Cu_A)-centre, with a carboxy-terminal fusion to a cytochrome *c* domain. The arrangement of the active site and redox centres is shown in Fig. 1b. In the context of number of amino acids and structured waters modelled, molecule 1 is the most complete and is the focus of the discussion that follows.

The core subunit, SU I/III, of caa_3-oxidase shares a distinct sequence and structure homology with its type A1 counterparts (Supplementary Fig. 1). Akin to ba_3-oxidase from *T. thermophilus*, the *a*-type haems in caa_3-oxidase have hydroxyethylgeranylgeranyl tails and are designated haem a_s[18]. Both haems a_s and a_{s3} are located in the hydrophobic core of the SU I domain (Fig. 1b) with their metal centres approximately 11 Å from the periplasmic surface. As in other HCOs, two histidines, His 73 and 387, act as axial ligands to the low-spin haem a_s iron at a distance of 2.0 and 2.1 Å, respectively (Supplementary Fig. 8a). Ligating the high-spin haem a_{s3} iron at its proximal side is a highly conserved His 385. The Cu_B is coordinated by the three canonical histidines, His 250, 299 and 300, and is located 4.9 Å from the haem a_{s3} iron (Supplementary Fig. 8b). The ligands bridging Cu_B and the high-spin iron have been modelled as a hydroxide and water pair 2.3 Å apart, with the hydroxide ion located 2.2 Å from the copper and the water 2.2 Å from the haem iron (Fig. 2a, b). Ligand assignments in the active site and oxygen entry into the enzyme (Fig. 2d and Supplementary Fig. 9) are discussed in the Supplementary Information.

SU I/III contains two proton channels, the D- and K-pathways (Fig. 3a, c). These are conduits for the protons that are (1) pumped across the membrane and (2) required for the chemical reduction of oxygen in the binuclear centre. A fuller description and comparison of the proton channels and exit pathways are included in the Supplementary Information (Supplementary Discussion and Supplementary Figs 10–12). The end of the D-pathway in caa_3-oxidase is distinctly different from the proton gating site of type A1 HCOs (Fig. 3b and Supplementary Fig. 13). The canonical type A1 gating glutamate, Glu 278 (*Paracoccus denitrificans* numbering), is replaced by Thr 252 in caa_3-oxidase whereas the phenylalanine–glycine pair, a single turn upstream of the type A1 glutamate position in helix VI, is replaced by a consecutive tyrosine (Tyr 248) and serine (Ser 249) residues (Supplementary Fig. 13). This YS pair (Supplementary Fig. 13b) offers a functional equivalent to the gating glutamate as determined by single-, double- and triple-mutant studies performed on the *P. denitrificans* aa_3-type oxidase mimicking the type A2 pump configuration in both *T. thermophilus* and *Rhodothermus marinus*[19]. In the caa_3-oxidase, Tyr 248 is held in position through interactions with the side chain amide of Asn 205 and the hydroxyl of Thr 252, whereas a hydrogen bond between Tyr 248 and a structured water (Wat 2074) connects the tyrosine to the D-pathway water chain (Fig. 3a, b).

Molecular dynamics simulations (MDS)[19] that accompanied the above mutagenesis work predicted a chain of four water molecules linking the engineered tyrosine to the active site. The water chain

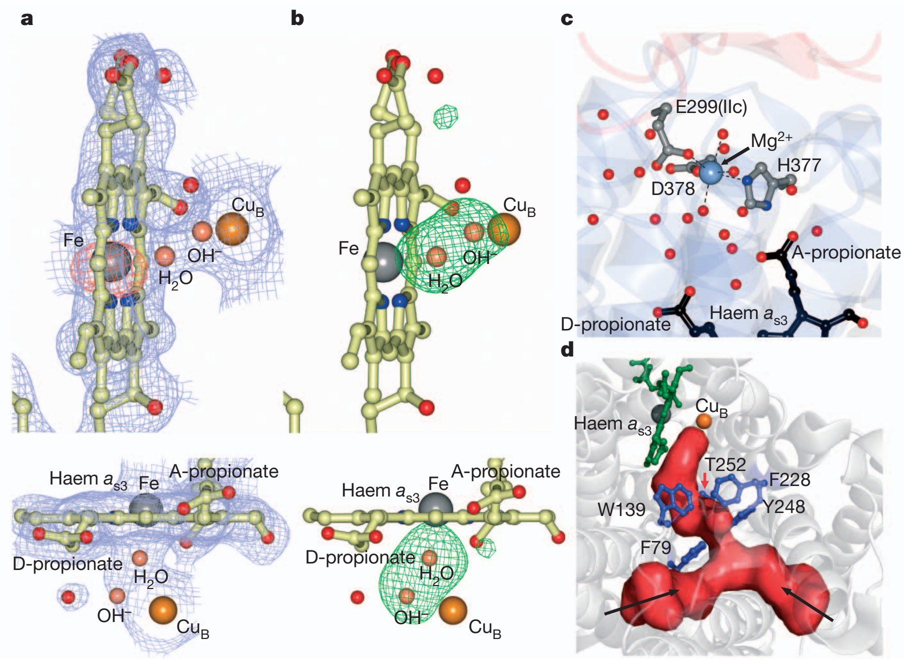


Figure 2 | Active site, water pool and oxygen channel in caa_3-oxidase. **a**, **b**, The active site of haem a_{s3} viewed from the oxygen entry side (top) and the periplasm (bottom). **a**, $2mF_o - DF_c$ electron density map contoured to 1σ (blue) and an anomalous difference map contoured to 5σ (red). **b**, $mF_o - DF_c$ electron density difference map calculated without the bridging ligands contoured to 4σ (green). For clarity, the anomalous map is not shown in **a** (bottom). **c**, Water pool. A water-filled cavity centred on a magnesium ion is situated above haem a_{s3}. Coordinate bonds are shown as dashed black lines. **d**, The calculated oxygen channel is Y-shaped and hydrophobic. Two extremities of the channel (red) contact the apolar surface of the protein (black arrows); a third contacts the binuclear centre.

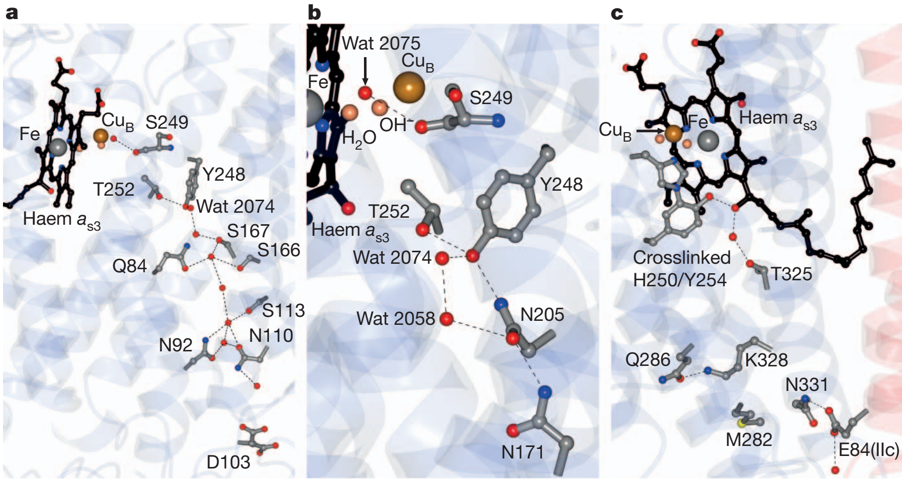


Figure 3 | Proton pathways in *caa*$_3$-oxidase. **a**, The D-pathway begins at Asp 103 and leads through a solvent-filled cavity to Tyr 248. Alternative conformations for the Asp 103 side chain were modelled into electron density and are shown. The conformer pointing towards Asn 110 represents the conformation observed in other known oxidases. **b**, Detailed view of the YS gate, the structural equivalent in *caa*$_3$-oxidase of the classical type A1 gating glutamate. **c**, The K-pathway originates at Glu 84 of SU IIc and continues up to the binuclear centre via the cross-linked His 250–Tyr 254, by way of Lys 328. Hydrogen bonds are shown as dashed black lines in **a**–**c**. The known covalent linkage between His 250 and Tyr 254 was modelled on the basis of separate biochemical characterization[32].

was anchored by the Nε2 of His 326, the hydroxyl of Tyr 274 and the carbonyl oxygen of Ser 275 (*P. denitrificans* numbering). The side chain of Ser 275, however, did not interact with the water chain. Similarly, in *caa*$_3$-oxidase, Wat 2075 links through a hydrogen bond to the carbonyl of Ser 249, whose side chain points away from the active site (Fig. 3b). Given that Thr 252 is not conserved in type A2 oxidases, we speculate that it functions as part of a protonic chain from Tyr 248 to the active site replacing the water that links through a hydrogen bond to the tyrosine hydroxyl in the aforementioned MDS studies. These observations lend support to the presence of a similar water chain in *caa*$_3$-oxidase which is only partially resolved in the current structure (Fig. 3b). MDS, carried out on the model reported here, identify transient water chains in the YS gate region that provide clear protonic connectivity to the binuclear centre and to the D-propionate of haem a_{s3} (Supplementary Fig. 11a, b).

In *caa*$_3$-oxidase from *R. marinus*, the tyrosine of the YS gate has been reported to deprotonate during the gating event as evidenced by Fourier transform infrared (FTIR) difference spectroscopy[20]. This observation is counter to earlier pK_a calculations, which indicate that the tyrosine remains protonated[21]. The pK_a of a tyrosine hydroxyl in aqueous solution is approximately 10. Given the hydrophobic environment surrounding Tyr 248 in the *caa*$_3$-oxidase of *T. thermophilus*, the effective pK_a of its hydroxyl is expected to be above 10. This has been confirmed by pK_a calculations performed on the *caa*$_3$ oxidase model reported here (Supplementary Table 2). The higher pK_a favours protonation, possibly limiting the role of Tyr 248 to one of stabilizing the water network in the proton pathway.

Subunit IIc is a fusion between the classical SU II and a cytochrome *c* (Fig. 1a). The SU II domain consists of two amino-terminal transmembrane helices anchoring a cupredoxin-like domain composed of a 10-stranded β-barrel that hosts the Cu$_A$ homobinuclear centre, the primary acceptor of electrons from cytochrome *c* (Fig. 4a). The asymmetric ligands for the Cu$_A$ centre in SU IIc are highly conserved among type A and B cytochrome *c* oxidases. The cytochrome *c* domain of SU IIc consists of four α-helices connected by random coil segments (Fig. 4a, Supplementary Discussion and Supplementary Fig. 14). The haem *c* resides within the hydrophobic core of the domain and is held in place by the classical axial iron ligands, His 251 and Met 303, and by two thioether linkages to Cys 247 and Cys 250.

Cytochrome *c* is oriented at the face of the cupredoxin domain such that the D-propionate of haem *c* is pointing towards the Cu$_A$ centre (Fig. 4a, b). This arrangement is consistent with speculation regarding the assembly of *caa*$_3$-oxidase in *R. marinus*[22,23]. It is, however, counter to that described on the basis of MDS and nuclear magnetic resonance studies of the interaction between cytochrome *c* and the cupredoxin domains of *ba*$_3$- and *aa*$_3$-oxidase, respectively, in solution[2,24,25]. Here, contact was proposed by way of the exposed haem cleft of the type reported for the cytochrome *c*/*bc*$_1$ complex from *Saccharomyces cerevisiae*[26]. Interestingly, the assembly of the cytochrome *c*/cupredoxin domain in *caa*$_3$-oxidase is quite distinct from the arrangement of the mono- and di-haem domains of *cbb*$_3$-oxidase (Supplementary Fig. 15).

The cytochrome *c* (SU IIc residues 236–325) and cupredoxin (SU IIc residues 117–216) domains share an interface with a calculated buried surface area of 934 Å^2 per domain (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Interaction between the domains is by surface and electrostatic complementarity (Fig. 4c). The contact surface of the cupredoxin domain is anionic and complementary to the cationic contact surface of the cytochrome *c* domain created by Lys 298 and Arg 272 in the vicinity of the haem propionates (Fig. 4c). Interdomain contacts are tabulated in Supplementary Table 3.

The optimal electron tunnelling pathway between haem *c* and Cu$_A$ in *caa*$_3$-oxidase was calculated using the program HARLEM (http://harlem.chem.cmu.edu) (Fig. 4b). It proceeds through the D-pyrrole and D-propionate of haem *c*, the backbone *cis*-amide nitrogen and α-carbon of Phe 126(IIc) and the sulphur of Met 208(IIc). The effective tunnelling length of the pathway is 18.8 Å with a direct metal-to-metal separation of 17.2 Å. By contrast, the closest 'edge-to-edge' distance between the haem *c* D-propionate donor and the Cu$_A$ centre acceptor moieties is 9.2 Å which too can be bridged efficiently by a tunnelling electron[27]. Such a process would involve two through-space jumps; the first from the D-propionate of haem *c* to the phenyl ring of Phe 126(IIc) (separation, 4.4 Å) and a second to either the sulphur of Cys 201(IIc) (separation, 3.7 Å) or directly to the Cu$_A$ centre (separation, 5.1 Å). Interestingly, mutating Trp 121 of the *P. denitrificans aa*$_3$-oxidase, a structural homologue of Phe 126(IIc), to a glutamine inactivates the enzyme[28–30]. Furthermore, changing the homologous Phe 88 in *ba*$_3$-oxidase to leucine resulted in a drop to 68% in electron transfer efficiency[2]. Together, these data indicate that the bulky hydrophobic side chain of Phe 126(IIc) is required to stabilize the conserved *cis*-amide for efficient electron transfer. A mechanism for electron entry into the cytochrome *c* domain is proposed (Supplementary Information).

The structure of *caa*$_3$-oxidase as described here offers the first view of a type A2 HCO consisting of a pair of fused canonical subunits and a

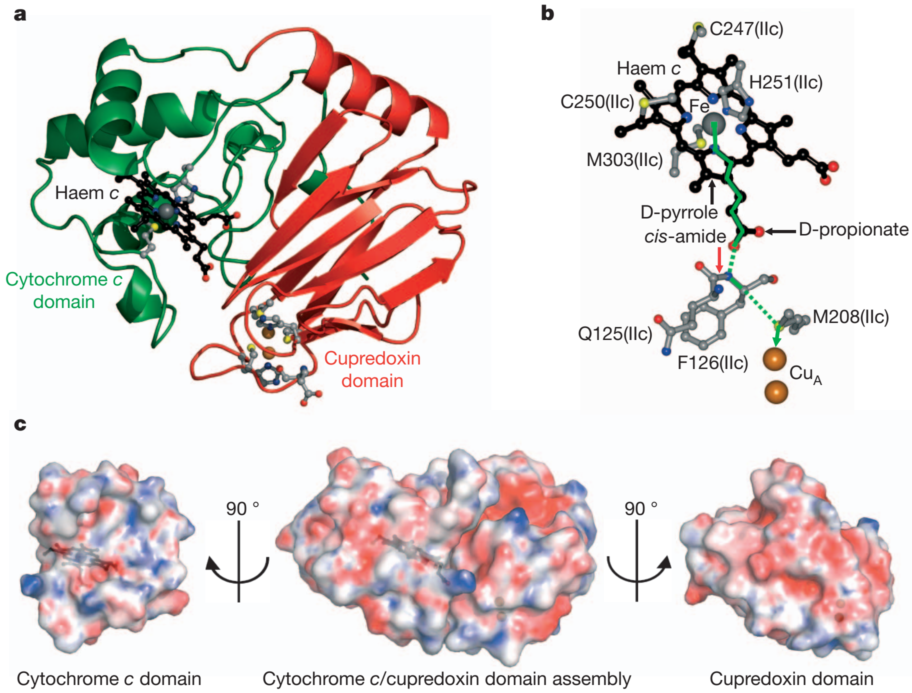


Figure 4 | Cytochrome *c*/cupredoxin complex and electron transfer pathway in *caa*$_3$-oxidase. a, Structure of the cytochrome *c* (green) and cupredoxin (red) domains of SU IIc. **b**, The optimum electron transfer pathway between the haem *c* iron and the Cu_A centre was calculated with HARLEM. Through bond and through space tunnelling are shown in solid and dashed green lines, respectively. **c**, 'Open book' view of surface complementarities between the cytochrome *c* and cupredoxin domains in *caa*$_3$-oxidase. Electrostatic potentials are viewed at $\pm$ 10 k_bT. The linker (residues 217–235) between the two domains has been removed for clarity. Surfaces are at 25% transparency to reveal the positions of haem *c* and the Cu_A centre shown in black ball and stick and copper spheres, respectively.

previously unknown SU IV. It shows the cytochrome *c*/oxidase complex where the D-propionate of haem *c* is the most probable conduit for electron-loading into Cu_A through the *cis*-amide of an aromatic residue conserved across type A and B cytochrome *c* oxidases. Electrons are proposed to enter the cytochrome *c* domain by way of its exposed haem edge. MDS calculations on the type A2 pump-gating site lend support to protonic water chains linking Tyr 248 to both the binuclear centre and the D-propionate of haem a_{s3}. Further, interaction between Lys 328 and Gln 286 is proposed to stabilize the canonical lysine in its protonated state in the K-pathway, replacing the functionally equivalent serine-water-lysine motif in type A1 HCOs.

The *caa*$_3$-oxidase structure provides a framework for examining how this integral membrane complex couples energy transduction to the complete reduction of oxygen. The manner in which type A2 oxidases direct protons across the membrane and electrons and protons to the active site of the complex can now be studied in detail as a result of having available the three-dimensional arrangement of key residues, metal ions and structured waters. Crystal structures of representatives from all HCO families have now been solved and the stage is set for a more complete understanding of how these intricate proteinaceous machines, with diverse oxygen affinities, have evolved.

METHODS SUMMARY

Procedures for the isolation and purification of *caa*$_3$-type cytochrome *c* oxidase from native *T. thermophilus* membranes are described in Methods. Crystals in the *I*2 space group were grown in glass sandwich plates at 20 °C by the *in meso* method using 7.7 MAG as the host lipid. The precipitant contained 14–21% (v/v) polyethyleneglycol 400, 0.1 M NaCl, 0–0.1 M Li_2SO_4 and 0.1 M sodium citrate pH 4.5–5.0. After 7–10 days of growth, crystals were collected and cryo-cooled directly in liquid nitrogen.

Diffraction data were collected at beamline 23-ID-D (GM/CA-CAT) of the Advanced Photon Source, beamline I24 of the Diamond Light Source and beamline PX1 of the Swiss Light Source using minibeam technology. Data were recorded in 1° oscillations with 1–2 s exposures. The complete data set comprised of data merged from a complete low-resolution (3.5 Å) pass and twenty high-resolution (up to 1.9 Å) 10–20° wedges. Full details of diffraction data collection and processing, structure solution and refinement and of MDS calculations are described in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 August 2011; accepted 4 May 2012.
Published online 1 July 2012.

1. Ferguson-Miller, S. & Babcock, G. T. Heme/copper terminal oxidases. *Chem. Rev.* **96,** 2889–2908 (1996).
2. Muresanu, L. *et al.* The electron transfer complex between cytochrome c_{552} and the Cu_A domain of the *Thermus thermophilus ba*$_3$ oxidase. A combined NMR and computational approach. *J. Biol. Chem.* **281,** 14503–14513 (2006).
3. Maneg, O., Ludwig, B. & Malatesta, F. Different interaction modes of two cytochrome-c oxidase soluble Cu_A fragments with their substrates. *J. Biol. Chem.* **278,** 46734–46740 (2003).
4. Maneg, O., Malatesta, F., Ludwig, B. & Drosou, V. Interaction of cytochrome *c* with cytochrome oxidase: two different docking scenarios. *Biochim. Biophys. Acta* **1655,** 274–281 (2004).
5. Pereira, M. M., Santana, M. & Teixeira, M. A novel scenario for the evolution of haem-copper oxygen reductases. *Biochim. Biophys. Acta* **1505,** 185–208 (2001).
6. Pereira, M. M., Sousa, F. L., Verissimo, A. F. & Teixeira, M. Looking for the minimum common denominator in haem-copper oxygen reductases: towards a unified catalytic mechanism. *Biochim. Biophys. Acta* **1777,** 929–934 (2008).
7. Misquitta, L. V. *et al.* Membrane protein crystallization in lipidic mesophases with tailored bilayers. *Structure* **12,** 2113–2124 (2004).
8. Caffrey, M., Lyons, J., Smyth, T. & Hart, D. J. in *Current Topics in Membranes* Vol. 63 (ed. L. J. DeLucas) Ch. 4 83–108 (Academic Press, 2009).
9. Höfer, N., Aragão, D., Lyons, J. A. & Caffrey, M. Membrane protein crystallization in lipidic mesophases. Hosting lipid effects on the crystallization and structure of a transmembrane peptide. *Cryst. Growth Des.* **11,** 1182–1192 (2011).
10. Li, D., Lee, J. & Caffrey, M. Crystallizing membrane proteins in lipidic mesophases. A host lipid screen. *Cryst. Growth Des.* **11,** 530–537 (2011).
11. Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477,** 549–555 (2011).
12. Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* **38,** 29–51 (2009).
13. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
14. Fee, J. A., Choc, M. G., Findling, K. L., Lorence, R. & Yoshida, T. Properties of a copper-containing cytochrome c_1aa_3 complex: a terminal oxidase of the extreme

thermophile *Thermus thermophilus* HB8. *Proc. Natl Acad. Sci. USA* **77,** 147–151 (1980).
15. Steffens, G. C., Biewald, R. & Buse, G. Cytochrome *c* oxidase is a three-copper, two-heme-A protein. *Eur. J. Biochem.* **164,** 295–300 (1987).
16. Mather, M. W., Springer, P., Hensel, S., Buse, G. & Fee, J. A. Cytochrome oxidase genes from *Thermus thermophilus*. Nucleotide sequence of the fused gene and analysis of the deduced primary structures for subunits I and III of cytochrome caa_3. *J. Biol. Chem.* **268,** 5395–5408 (1993).
17. Mather, M. W., Springer, P. & Fee, J. A. Cytochrome oxidase genes from *Thermus thermophilus*. Nucleotide sequence and analysis of the deduced primary structure of subunit IIc of cytochrome caa_3. *J. Biol. Chem.* **266,** 5025–5035 (1991).
18. Lübben, M. & Morand, K. Novel prenylated hemes as cofactors of cytochrome oxidases. Archaea have modified hemes A and O. *J. Biol. Chem.* **269,** 21473–21479 (1994).
19. Backgren, C., Hummer, G., Wikstrom, M. & Puustinen, A. Proton translocation by cytochrome *c* oxidase can take place without the conserved glutamic acid in subunit I. *Biochemistry* **39,** 7863–7867 (2000).
20. Pereira, M. M., Sousa, F. L., Teixeira, M., Nyquist, R. M. & Heberle, J. A tyrosine residue deprotonates during oxygen reduction by the caa_3 reductase from *Rhodothermus marinus. FEBS Lett.* **580,** 1350–1354 (2006).
21. Soares, C. M., Baptista, A. M., Pereira, M. M. & Teixeira, M. Investigation of protonatable residues in *Rhodothermus marinus* caa_3 haem-copper oxygen reductase: comparison with *Paracoccus denitrificans* aa_3 haem-copper oxygen reductase. *J. Biol. Inorg. Chem.* **9,** 124–134 (2004).
22. Santana, M., Pereira, M. M., Elias, N. P., Soares, C. M. & Teixeira, M. Gene cluster of *Rhodothermus marinus* high-potential iron-sulfur protein: oxygen oxidoreductase, a $caa_{(3)}$-type oxidase belonging to the superfamily of heme-copper oxidases. *J. Bacteriol.* **183,** 687–699 (2001).
23. Srinivasan, V. *et al.* Structure at 1.3 Å resolution of *Rhodothermus marinus* caa_3 cytochrome *c* domain. *J. Mol. Biol.* **345,** 1047–1057 (2005).
24. Wienk, H. *et al.* Interaction of cytochrome *c* with cytochrome *c* oxidase: an NMR study on two soluble fragments derived from *Paracoccus denitrificans. Biochemistry* **42,** 6005–6012 (2003).
25. Roberts, V. A. & Pique, M. E. Definition of the interaction domain for cytochrome *c* on cytochrome *c* oxidase. *J. Biol. Chem.* **274,** 38051–38060 (1999).
26. Lange, C. & Hunte, C. Crystal structure of the yeast cytochrome bc_1 complex with its bound substrate cytochrome *c. Proc. Natl Acad. Sci. USA* **99,** 2800–2805 (2002).
27. Page, C. C., Moser, C. C. & Dutton, P. L. Mechanism for electron transfer within and between proteins. *Curr. Opin. Chem. Biol.* **7,** 551–556 (2003).
28. Drosou, V., Malatesta, F. & Ludwig, B. Mutations in the docking site for cytochrome *c* on the *Paracoccus* heme aa_3 oxidase. Electron entry and kinetic phases of the reaction. *Eur. J. Biochem.* **269,** 2980–2988 (2002).
29. Drosou, V., Reincke, B., Schneider, M. & Ludwig, B. Specificity of the interaction between the *Paracoccus denitrificans* oxidase and its substrate cytochrome *c*: comparing the mitochondrial to the homologous bacterial cytochrome c_{552}, and its truncated and site-directed mutants. *Biochemistry* **41,** 10629–10634 (2002).
30. Witt, H., Malatesta, F., Nicoletti, F., Brunori, M. & Ludwig, B. Tryptophan 121 of subunit II is the electron entry site to cytochrome-*c* oxidase in *Paracoccus denitrificans*. Involvement of a hydrophobic patch in the docking reaction. *J. Biol. Chem.* **273,** 5132–5136 (1998).
31. Lomize, A. L., Pogozheva, I. D., Lomize, M. A. & Mosberg, H. I. Positioning of proteins in membranes: a computational approach. *Protein Sci.* **15,** 1318–1333 (2006).
32. Buse, G., Soulimane, T., Dewor, M., Meyer, H. E. & Bluggel, M. Evidence for a copper-coordinated histidine-tyrosine cross-link in the active site of cytochrome oxidase. *Protein Sci.* **8,** 985–990 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from Science Foundation Ireland Grants 07/IN.1/B1836 (M.C.) and BICF685 (T.S.), the National Institutes of Health Grants GM75915, P50GM073210, U54GM094599 (M.C.), and FP7 COST CM0902 (M.C.) and Marie Curie Actions PIEF-GA-2009-235612 (D.A.). We thank D. Hart, J. Lee and T. Smyth for help with 7.7 MAG synthesis, M. Dewor for amino acid sequencing, C. Soares for discussions regarding surface potential calculations, and G. Winter for help with data reduction from multiple crystals. The assistance and support of beamline scientists at the Advanced Photon Source (23-ID), Diamond Light Source (I24) and Swiss Light Source (PX1) are gratefully acknowledged.

Author Contributions J.A.L. synthesized 7.7 MAG, optimized crystallization conditions, collected and processed data, solved, refined and analysed the structure, and wrote the manuscript. D.A. collected crystals, collected and processed data, solved, refined and analysed the structure, and wrote the manuscript. O.S. performed initial protein production, purification, crystallization and data collection. A.V.P. carried out molecular dynamics simulations and pK_a calculations. T.S. initiated the project, produced, purified and characterized protein, helped with crystallization, data collection and processing and structure analysis, and wrote the manuscript. M.C. was responsible for the overall project strategy and management and oversaw manuscript preparation.

Author Information Coordinates and structure factors for caa_3-oxidase are deposited in the Protein Data Bank (accession code 2YEV). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.S. (tewfik.soulimane@ul.ie) or M.C. (martin.caffrey@tcd.ie).

METHODS

Protein production and purification. Biomass of the extreme thermophilic bacterium *T. thermophilus* HB8 was produced by fermentation at the Helmholtz Centre for Infection Research (Braunschweig, Germany), as described previously[33]. The protocol used for isolating and purifying caa_3-oxidase from the biomass is similar to that used for the ba_3-oxidase with slight modifications[33]. Briefly, 100 g of *T. thermophilus* biomass was suspended in 0.5 l 0.1 M Tris-HCl pH 7.6 buffer (buffer A) containing 0.2 M KCl and homogenized using a blender. Lysozyme was added to a final concentration of 0.6 μM and the suspension was stirred for 3 h at 4 °C. After centrifugation at 53,936*g* for 45 min at 4 °C the pellet was resuspended in 0.5 l buffer A, homogenized and centrifuged again at 53,936*g* for 30 min at 4 °C. This washing step was repeated three times. The final pellet was resuspended in 0.5 l buffer A and incubated with stirring for 3 h at 4 °C in the presence of 5% (v/v) Triton X-100 (Sigma Aldrich) to solubilize the membranes. After centrifugation at 53,936*g* for 1 h at 4 °C, the supernatant was collected and diluted with MilliQ water at 4 °C to a final volume of 5 l. It was subsequently loaded at 4 °C on DEAE-Biogel agarose anion exchange resin (Biorad) packed in a 30 cm × 10 cm column equilibrated with 0.01 M Tris-HCl pH 7.6, 0.1% (v/v) Triton X-100 (buffer B). After washing with 2 l buffer B, the respiratory complexes were eluted with 4 l of a 0 to 0.2 M linear gradient of NaCl in buffer B.

The caa_3-oxidase containing fractions, identified by ultraviolet/visible spectroscopy, from the anion exchange column were pooled (circa 0.8 l) and dialysed against 10 l buffer B for 5 h at 4 °C. The dialysed material was loaded on an XK 26/20 column packed with 30 ml Fractogel EMD TMAE anion exchange resin (Merck) pre-equilibrated with buffer B. With the protein bound to the resin, the Triton X-100 detergent was exchanged for 0.05% (w/v) *n*-dodecyl-β-D-maltoside (DDM). The protein was eluted with a 0 to 0.3 M linear gradient of NaCl in 0.01 M Tris-HCl pH 7.6, 0.05% (w/v) DDM (buffer C) for 1 h at 22 °C and at a flow rate of 4 ml min^{-1}. The caa_3-oxidase containing fractions (circa 0.16 l) were pooled and concentrated to 8 ml using a centrifugal filter (Centricon YM100, Millipore) and applied to a Highload XK 16/60 Superdex 200 gel filtration column (GE-Healthcare) previously equilibrated with 0.05 M Tris-HCl pH 7.6, 0.05% (w/v) DDM. Fractions across the caa_3-oxidase peak were pooled (circa 90 ml) and diluted tenfold with 0.01 M sodium phosphate pH 6.8 containing 0.05% (w/v) DDM, and applied to an XK 16/20 column packed with 20 ml hydroxylapatite 'high resolution' ion exchange resin (Fluka, BioChemika) pre-equilibrated with the same buffer. The caa_3-oxidase was eluted with a 0.01 to 0.04 M linear gradient of sodium phosphate pH 6.8 containing 0.05% (w/v) DDM within 0.5 h at a flow rate 2 ml min^{-1} at 22 °C. The pooled caa_3-oxidase fractions (circa 50 ml) were diluted fivefold with buffer C reducing the sample conductivity to < 2 mS cm^{-1}. The previous anion exchange and concentration steps using buffer C was repeated once. The detergent was exchanged to 0.2% (w/v) *n*-decyl-β-D-maltoside in 0.01 M Tris-HCl pH 7.6 and 150 mM NaCl during a final Highload XK 16/60 Superdex 200 gel filtration step and resulted in circa 10 mg of purified caa_3-oxidase. The protein was concentrated to 10 mg ml^{-1} (Centricon YM100, Millipore), flash-frozen in liquid nitrogen and stored in aliquots of 20 μl at −80 °C. The protein was used in the 'as-prepared' oxidized state for crystallization studies.

Crystallization. 7.7 MAG was synthesized according to ref. 8. The protein-laden mesophase was prepared by homogenizing 7.7 MAG and a 10 mg ml^{-1} protein solution, supplemented with sodium ascorbate to a final concentration of 0.2 μM, in a 1:1 ratio by weight using a dual syringe mixing device at 20 °C[34]. Crystallization trials were carried out at 20 °C in 96-well glass sandwich plates with 50 nl mesophase and 0.8 μl precipitant solution using an *in meso* robot[35]. Precipitant solutions consisted of 14–21% (v/v) PEG 400, 0.1 M NaCl, 0–0.1 M Li_2SO_4 and 0.1 M sodium citrate pH 4.5–5.0. Triangular plate-like crystals grew to a maximum size of 100 × 60 × 5 μm^3 in 7 to 10 days (Supplementary Fig. 2). Wells were opened using a tungsten-carbide glass cutter and the crystals were harvested using 100-μm micromounts (MiTeGen). Crystals were cryo-cooled directly in liquid nitrogen.

Subunit IV identification and sequencing. The caa_3-oxidase complex, pre-purified by vapour diffusion crystallization using a precipitant of 10 mM *bis*-Tris pH 6.0, 6% (w/v) PEG 2000 and 100 mM NaCl, was separated into its individual subunits on a reversed phase Synchropak C_4 column (250 mm × 4.6 mm; Supelco) using a Hewlett-Packard (HP) 1050 HPLC system at 40 °C. The solvents used were (A) 5% formic acid, (B) 95% formic acid, (C) n-propanol, and (D) acetonitrile. Gradients were formed with a microprocessor-controlled quaternary pump (HP) by low-pressure mixing of the four solvents. Two solvent gradients were used. Gradient 1 used (A) 40–5%, (B) constant at 60%, (C) 0–10%, and (D) 0–25% in the 0–30 min run time. Gradient 2 used (A) constant at 5%, (B) constant at 60%, (C) 10–25%, and (D) 25–10% in the 30–60 min run time. Separated subunits were identified by amino-terminal sequencing after deformylation as appropriate. Parenthetically, we note that vapour diffusion grown crystals, referred to above, diffracted to no better than 3.8 Å and were unsuitable for structure determination.

The full length SU IV was characterized by amino- (residues 1–41) and carboxy-terminal (residues 60–66) sequencing and mass spectrometry. Sequencing was performed on peptide fragments prepared using cyanogen bromide, as reported for SU IIa in ba_3-oxidase[36]. The peptide fragments were separated by HPLC at 45 °C, as described above. The solvents used were (A) 70% formic acid and (B) 70% formic acid, 15% n-propanol and 11% acetonitrile with a linear gradient of 100%A to 100%B. Sequencing results are shown in Supplementary Fig. 5.

Data collection and processing. X-ray diffraction data were collected on the 23-ID-D beamline of the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) at the Advanced Photon Source (APS), Argonne, Illinois, USA, the I24 beamline at the Diamond Light Source (DLS), Didcot, Oxford, UK, and the PX1 beamline at the Swiss Light Source (SLS), Villigen, Switzerland. Data were acquired using a 10-μm collimated minibeam at GM/CA-CAT[37], whereas a 10-μm and a 20-μm microfocus beam was used at DLS and SLS, respectively. Attenuated (10×) images were used to locate and to centre on highly ordered regions of the crystal[38]. Oscillation data were measured in 1.0° frames with 1–2 s exposures using a 1× or 10× attenuated beam. All data were initially reduced in xia2[39] using XDS[40], XSCALE and SCALA[41]. Optimum data wedges were identified by data quality and isomorphic unit cell parameters. These data were rescaled in XSCALE and merged in SCALA[41] (Supplementary Table 1). The data reduction strategy involved combining a complete low resolution (3.5 Å) data set recorded from a single crystal with twenty high resolution (up to 1.9 Å) 10–20° wedges of data collected from multiple crystals.

Structure solution and refinement. Molecular replacement search models were prepared from SU I, SU II and SU III of the bovine heart cytochrome *c* oxidase complex (PDB ID: 2OCC) pruned of all non-protein atoms using Chainsaw in CCP4[42]. Initial phases were obtained by molecular replacement as performed by the program Phaser[43], identifying two molecules (molecules 1 and 2) in the asymmetric unit. Initial stages of structure refinement and model building were performed after simulated annealing in CNS[44] with subsequent rounds of refinement carried out in PHENIX[45]. NCS restraints were used throughout and the inclusion of TLS refinement (20 groups over molecules 1 and 2) was implemented in the latter stages of refinement. The structural model was revised in real space with the program COOT[46] using sigma-A weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps. The geometric quality of the model was assessed with MolProbity[13]. Structures of cofactors, lipids and water molecules were determined and refined as described below.

Cofactors. The locations of two haems a_s, haem *c*, the binuclear Cu_A, Cu_B and magnesium in each molecule in the asymmetric unit were determined using the $2F_o - F_c$ electron density map. Iron atoms of the three haems were located using an anomalous difference Fourier map from data collected at the iron K-edge (1.7397 Å) at the SLS.

Lipids and unassigned density. Three types of lipid molecule were identified in a composite omit $2F_o - F_c$ map. The first, a native *Thermus* lipid, was found buried within the protein at the interface between the SU I and SU III domains of the fused SU I/III. The second includes the 1- and 2- MAG isoforms of 7.7 MAG, where 1- and 2- refers to the position of the acyl chain on the glycerol headgroup. The third was modelled as a diacylglycerol with two fatty acyl chains. The diacylglycerol may derive from the hosting 7.7 MAG by way of transesterification or from a more complex native lipid. Electron density in the vicinity of the headgroup region of the diacylglycerol was ill-defined. Tentatively modelled lipid and solvent molecules were removed if no clear hydrogen bonding was present and/or the model was not justified by the electron density.

Water molecules. Both surface associated and integral water molecules were assigned based on sigma-A weighted $2F_o - F_c$ electron density maps contoured at 1σ using standard geometrical and chemical restraints. Molecules 1 and 2 have 247 and 145 waters, respectively, for a total of 392 water molecules per asymmetric unit.

Electrostatic surfaces. The electrostatic potential was calculated separately on the cupredoxin domain (residues 117–216) and the cytochrome *c* domain (residues 236–325) of SU IIc using the Poisson–Boltzmann method as implemented in the Potential module from the package MEAD[47]. The ionic strength in the calculations was 0.1 M. A grid of 141 points with 1-nm spacing was used to enclose the protein. The solvent probe radius was set to 1.4 Å and the ion exclusion layer thickness was 2.0 Å. An external dielectric constant (ε) of 80 was used for the solvent and the internal ε was set to 4. PARSE atomic charges and radii were used. The reduced haem *c* iron was assigned a charge of +1.4 with the remaining +0.6 charge assigned equally amongst its axial ligands. The oxidized Cu_A site was assigned a mixed valence of +1.0 on each copper ion with the remaining +1.0 charge assigned proportionally based on coordination to its axial ligands.

MD simulations. An all-atom simulation system included caa_3-oxidase embedded into a pre-equilibrated POPE lipid bilayer solvated in 0.15 M NaCl. All crystallographic water molecules were kept and additional water molecules were modelled in the protein internal hydrophilic cavities using Dowser[48]. The total simulation system size was ~159,000 atoms. MDS were performed using NAMD[49] with the CHARMM force field for proteins (C22 with the CMAP correction)[50,51], lipids (C36)[52] and water (the TIP3P water model). A lipid molecule identified in SU I/III (Supplementary Fig. 6) was parameterized using the CHARMM General Force Field[53]. All protein residues were modelled in their standard protonation states and metal centres were modelled as in the reduced state. Production runs (two independent simulations) were performed in the NPT ensemble at a constant temperature (310 K) and pressure (1 atm) for the combined simulation time of ~300 ns. Coordinates were saved every 2 ps. VMD[54] was used for trajectory analysis.

pK_a calculations. Electrostatic calculations were performed using MEAD[47] with the following options: protein atomic coordinates were taken from the X-ray structure; hydrogen atoms were added and minimized. MEAD requires that the dielectric constant used for the protein is also used for the membrane. Accordingly the pK_a calculations were carried out using $\varepsilon = 4$ and 10 for the protein and membrane regions and $\varepsilon = 80$ for solvent. Calculations were performed, and results are reported, with and without explicit water molecules.

Electron transfer. The electron transfer pathway calculation between the haem *c* iron of the cytochrome *c* domain and homobinuclear Cu_A centre, with both sites selected interactively, was performed using the pathways module as implemented in HARLEM. The non-bonding and hydrogen bonding decay was 1.7 $Å^{-1}$ and the covalent bond decay was 0.6 $Å^{-1}$. The choice of the best electron transfer pathway was based on couplings.

Oxygen channels. Oxygen channel cavities were calculated with HOLLOW[55] using a grid spacing of 0.25 Å and a probe radius of 1.4 Å.

Structure-based sequence alignment. Coordinates for the type A cytochrome *c* oxidases present in the PDB (PDB IDs: 3HB3, 2GSM, 1V54) were aligned with the caa_3-oxidase structure, presented here (PDB ID: 2YEV), using the secondary structure matching algorithm (SSM)[56] as implemented in COOT[46]. This initial alignment was further refined with Modeller[57] with a threshold of 3.5 Å between C_α atoms for inter-protein residue matching. Figures for the resulting structure-based sequence alignments were produced with ALINE[58]. Percentage sequence identity was calculated from the number of identical residues/minimum length.

Protein figures were generated in CCP4mg. Figures describing oxygen cavities and electrostatic surfaces were generated in PyMOL.

33. Soulimane, T., Kiefersauer, R. & Than, M. E. in *Membrane Protein Purification and Crystallization (Second Edition)* (eds Hunte, C., von Jagow, G. & Schägger, H.) Ch. 14 229–251 (Academic Press, 2003).
34. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4,** 706–731 (2009).
35. Cherezov, V., Peddi, A., Muthusubramaniam, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60,** 1795–1807 (2004).
36. Soulimane, T., Than, M. E., Dewor, M., Huber, R. & Buse, G. Primary structure of a novel subunit in ba_3-cytochrome oxidase from *Thermus thermophilus. Protein Sci.* **9,** 2068–2073 (2000).
37. Fischetti, R. F. *et al.* Mini-beam collimator enables microcrystallography experiments on standard beamlines. *J. Synchrotron Radiat.* **16,** 217–225 (2009).
38. Cherezov, V. *et al.* Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 micron size X-ray synchrotron beam. *J. R. Soc. Interface* **6,** S587–S597 (2009).
39. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43,** 186–190 (2010).
40. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
41. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).
42. Collaborative Computational Project, number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
43. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
44. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
45. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
47. Bashford, D. & Gerwert, K. Electrostatic calculations of the p*K*a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **224,** 473–486 (1992).
48. Zhang, L. & Hermans, J. Hydrophilicity of cavities in proteins. *Proteins* **24,** 433–438 (1996).
49. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
50. MacKerell, A. D. *et al.* All atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
51. MacKerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).
52. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114,** 7830–7843 (2010).
53. Vanommeslaeghe, K. *et al.* CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31,** 671–690 (2010).
54. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
55. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8,** 49 (2008).
56. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60,** 2256–2268 (2004).
57. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
58. Bond, C. S. & Schuttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D* **65,** 510–512 (2009).